Essays on the Political Economy of Development

Qing Zhang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

Essays on the Political Economy of Development

Qing Zhang

My dissertation studies political economy issues in the development of China. Chapter 1 lies in the intersection of urban economics and political economy. This chapter exploits plausibly exogenous variation generated by a unique national policy in China that requires all residential buildings to receive sufficient hours of sunshine to study a central question in urban economics, namely, whether urban density facilitates the diffusion of information. The policy creates higher degrees of restriction on density at higher latitudes, where longer shadows require buildings to be further apart. Data on individual housing projects across China reveal that the cross-latitude variation in regulatory residential Floor Area Ratio can be described quite well by a formula linking structure density to latitude through the solar elevation angle. These differences in building density further induce differences in population density and land prices across latitudes. Using differential topic dynamics on a national petition platform to measure information diffusion, this chapter shows that people respond to shifts in government attention with varying speeds across latitudes. Increases in local government reply rate to a topic raises the volume of subsequent posts on the same topic, exhibiting an S-shaped time trajectory consistent with local information diffusion about shifting government priorities. These responses are systematically faster in southern cities, where density is higher. Survey evidence further indicates that otherwise similar individuals are more likely to gossip about public issues in a southern city.

Chapter 2, coauthored with Junyan Jiang and Tianguang Meng and forthcoming at the journal *Governance*, examines the flip side of the interaction between local governments and citizens studied in Chapter 1. This chapter studies the response of government policies to opinions expressed online. We address this question by studying the patterns and consequences of online participation at a major electronic petition platform in China. Content analysis of around 900,000 petitions reveals that a substantial share of them con-

cern lower-class issues and are originated from less developed rural and suburban areas. Linking variations in petition volumes to an original dataset of government policy priorities, we further show that online participation led governments to place greater emphasis on social welfare policies and to increase the coverage of a key low-income assistance program. These results underscore the potential of online participation as an important mechanism to improve the quality of governance.

Chapter 3, coauthored with Amit Khandelwal, Suresh Naidu and Heiwai Tang, turns to a systematic examination of China's reform process. We apply natural language processing methods to analyze a comprehensive corpus of 1.4 million legal documents issued by the Chinese government at central and local levels since 1949, and measure their market orientation in a data-driven fashion. We document an active introduction of market-oriented legal infrastructure from the mid-1980s to around 2000, which slowed down in the last fifteen years. These dynamics are present within fine-grained policy domains. We find that the market orientation of policies explains just an extra 2% of provincial variation in GDP per capita growth beyond province and time fixed effects. Variable selection based on richer representations of the text exhibits similarly limited predictive power for provincial growth. Taken together, these findings suggest the importance of studying the informal arrangements between market participants and government officials.

Contents

List of Figures **iv**

List of Tables **vi**

Acknowledgements **viii**

1 Sunlight, Urban Density and Information Diffusion **1**

1.1 Introduction 1

1.2 Institutional Background 8

1.3 Description of Data 12

1.4 Effects of the Policy on Building Height and FAR 14

1.5 Relationship between Density and the Speed of Information Diffusion . . 26

1.6 Threats to Validity 36

1.7 Conclusion 38

2 From Internet to Social Safety Net: The Policy Consequences of Online Participation in China **59**

2.1 Introduction 59

2.2 Background 62

2.3 Hypotheses 65

2.4 Analyzing LLMB Petitions: Issues and User Backgrounds 68

2.5 The Policy Consequences of Online Participation 70

2.6 Concluding Remarks 79

3 Do Legal Rules Explain China's Economic Growth? 83

3.1 Introduction 83

3.2 Data 85

3.3 Descriptive Statistics 87

3.4 Measuring Market Orientation in Policies 92

3.5 Do Legal Rules Explain Economic Growth? 101

3.6 Conclusion 108

Bibliography 110

A Appendix to Chapter 1 120

A.1 Details of Dataset Construction 120

A.2 Value of Lifting Regulatory FAR 122

A.3 Relationship between Regulatory FAR and Latitude, Raw Data 123

A.4 Density in Beijing, Shanghai and Shenzhen 124

A.5 Petition Topics 125

A.6 Longitude as a Placebo 128

B Appendix to Chapter 2 137

B.1 Electronic Government Petition Platforms around the World 137

B.2 Survey Evidence on Policy Preferences across Socioeconomic Status . . . 138

B.3 A Snap Shot of LLMB's Interface 139

B.4 Comparing Petition Activities on LLMB vs. Local Platforms 139

B.5 Details on Non-Parametric Content Analysis of Petition Issues and Users Backgrounds 141

B.6 Details on Estimating Government Policy Priorities 142

B.7 Geographic Distribution of Online Petitions 149

B.8 Details on Estimating and Classifying Online Petitions using LDA Models 150

B.9 Data on Collective Protests 154

B.10 Validating the Link between Petition and Rural Participants 155
B.11 Baseline Results with All Coefficients 157
B.12 Effect of Petition on All GWR Policy Topics 158
B.13 Robustness Checks 159
B.14 Granger-Style Tests on Reverse Causality 164
B.15 Instrumental Variables Analysis 166
B.16 Extensions 172
B.17 Fieldwork Evidence on the Petition-Policy Link 181

C Appendix to Chapter 3 **183**

C.1 Aggregate Trends for Washington Consensus Components 183

List of Figures

1.1 Sunlight Distance Coefficients and Latitude 53

1.3 Solar Elevation Angle and Building Height 53

1.5 Illustrations of Building Layout 54

1.7 High-rises in the North and the South, by Construction Year 54

1.9 Binned Scatterplot of Regulatory Floor Area Ratio and Latitude 55

1.11 Dynamic Response of Postings to a 10% Increase in Government Reply Rate 56

1.13 Cumulative Increase in Postings after a 10% Increase in Government Reply Rate 57

1.15 Cumulative Differential Responses of Postings to a 10% Increase in Reply Rate 58

2.1 Aggregated Distribution of GWR Topic Shares, 2000-2014 72

3.1 Number of Documents over Time 86

3.3 Screenshot of a Sample Document 87

3.5 Top Formal Law Categories 88

3.7 Share of Laws Issued by Local Governments 89

3.9 Share of Laws on Economic Issues 89

3.11 Share of Laws Effective in 2015 90

3.13 Share of Laws Still Effective n Years Since Issuance 91

3.15 Survival Curves for Economic vs. Non-Economic Laws 91

3.17 The Rise and Fall of "Commodity Economy" and "Market Economy" 94

3.19 Market Orientation in Legal Documents over Time 99

3.21 Washington Consensus over Time, Coastal vs. Non-Coastal Provinces 100
3.23 Trends of Market Orientation within Fine Areas of Law 102
3.25 Predictive Power of Legal Rules 109

A1 Identifying Urban Clusters 132
A3 Zipf's Law for Urban Clusters 133
A5 Relationship Between Latitude and Regulatory FAR Upper Limit 134
A7 Urban Density in Beijing, Shanghai and Shenzhen 135
A9 Urban Density in Beijing, Shanghai and Shenzhen, City Centers 136

B6 Effect of Petitions on All GWR Topics 158
B1 A Snapshot of the LLMB 174
B2 Posterior Distribution of Topic Proportions in the Whole Corpus 175
B3 Topic Quality: Coherence vs. Specificity 176
B4 Average Number of Petitions per Capita, 2008-2013 177
B5 Clear Dominant Topic in Petitions 177
B7 Level of LLMB Participation by Presence of Active Local Petition Platform 178
B8 Permutation of Topics Used for Instrument 178
B9 Effect of Online Participation by Year 179
B10 The Geography of Policy Responsiveness 180

C1 Washington Consensus Components in Formal vs. Informal Laws 184
C1 Washington Consensus Components in Formal vs. Informal Laws 185

List of Tables

1.1 Effects of the Sunlight Policy on Building Height 41

1.2 Parallel Economic Growth across Latitudes 42

1.3 Predictive Power of the Sunlight Policy 43

1.4 Is the Sunlight Policy Always a Binding Constraint on Density? 44

1.5 Latitude, Density and Prices 45

1.6 Reaction of Petition Posting to Government Replies 46

1.7 Dynamic Reaction of Petition Posting to Government Replies 47

1.8 Differential Speed of Petition Reaction to Government Replies 48

1.9 Differential Speed of Petition Reaction to Government Replies Using Observed FAR 49

1.10 Survey Evidence on Word of Mouth 50

1.11 Attitudes towards Government not Systematically Different Across Latitudes 51

1.12 Internet Use not Systematically Different Across Latitudes 52

2.1 Estimated Distributions for Issues and Petitioners' Backgrounds 69

2.2 Main Results 77

2.3 Change in Substantive Outcomes: Rural Minimum Living Standard Guarantee Scheme (*dibao*) 80

3.1 Areas of Law Experiencing Largest Changes in Share 92

3.2 Anchor Words and Their Relative Shares 96

3.3 Do Legal Rules Explain GDP Growth? 104

3.4 Do Legal Rules Explain FDI? 105
3.5 Provinces with Highest Shares of Bellwether Laws 107

A1 Data Sources 129
A2 Value of Loosening Regulatory FAR Upper Limit 130
A3 Petition Reaction to Government Replies, Longitude Placebo 131

B1 A List of Electronic Petition Platforms Created by National/Regional Governments 137
B2 Different Demands from High- and Low-Income Individuals 139
B3 Correlation in Petition Volumes: LLMB vs. Local Platforms 140
B4 Validating the Link between Online Petitions and Rural Participants 156
B5 Baseline Results with Coefficients for All Control Variables 157
B6 Baseline Estimates without Fixed-Effects 159
B7 Using Welfare Topic from 20-Topic LDA Model as the Dependent Variable . 160
B8 Sub-Sample Analysis: Including Low Petition Provinces 161
B9 Sub-Sample Analysis: By Region 162
B10 Sub-Sample Analysis: By Presence of Local Platforms 163
B11 Alternate the Timing of the Dependent Variable 164
B12 Controlling for Welfare Topic Share at the Higher Level or from Earlier Periods 165
B13 Instrumental Variable Estimation 167
B14 Change in Substantive Outcomes: Rural Minimum Living Standard Guarantee Scheme (*dibao*) 168
B15 UN Petition and Provincial Welfare Policy 169
B16 UN Petitions and A City's Own Welfare Outcomes 170
B17 Alternative Instruments 171
B18 Placebo using Urban Dibao as Outcome 172
B19 Effects of Online Petitions on Fiscal Expenditures 172

Acknowledgements

This dissertation would not have been completed without the unfailing support and guidance of my advisors, W. Bentley MacLeod, Suresh Naidu and Amit Khandelwal over the course of my graduate studies. Despite different research styles and interests, the three of them share a common dedication to the truth, a pursuit for answers to first-order questions, and a passion about understanding real-world institutions. I have been constantly challenged and inspired by them. As the range of topics that can be studied in an economics Ph.D. program becomes increasingly broad, it sometimes seems perplexing what constitutes economics. I am deeply grateful that my advisors have always encouraged me to study topics I am truly excited about, while constantly keeping me on the economist's leash.

My six years in graduate school were enriched by the amazing intellectual environment at Columbia. I benefited from discussions with many people, most notably Dan O'Flaherty, Jack Willis, Michael Best, Yeon-Koo Che, Don Davis, Andrea Prat, Bernard Salanié, Tobias Salz, Joseph Stiglitz, Eric Verhoogen and Shang-Jin Wei. I have also had the good fortune to exchange ideas with scholars from other institutions, including Alain Bertaud, Bruce Bueno de Mesquita, Junyan Jiang, Tianguang Meng, Mushfiq Mobarak, Enrico Moretti, David Ong, and Heiwai Tang.

I am also grateful to my fellow students at Columbia. We did it all together. In particular, I wish to extend my thanks to Yifeng Luo, Haaris Mateen, Anurag Singh, Yue Yu and Danyan Zha, without the support of whom I would have been substantially less productive and happy.

Finally, I thank my family and friends for their unfaltering trust. I dedicate this dissertation to my parents, who have had the ultimate patience to bear with their son while he was doing something seemingly useless in a continent thousands of miles from home.

To my parents

Chapter 1

Sunlight, Urban Density and Information Diffusion

1.1 Introduction

Density is the defining characteristic of cities. The exchange of information, knowledge and ideas facilitated by urban density is believed to result in higher productivity (Jacobs, 1969) and constitute an important source of urban agglomeration economies (Rosenthal and Strange, 2004). In the words of Marshall (1890), "The mysteries of the trade become no mysteries; but are as it were in the air."

However, empirically testing for the effect of density in accelerating information transmission is challenging due to both difficulties in measuring information diffusion and a paucity of exogenous sources of variation in density. On the one hand, information is typically unobservable to the researcher. On the other hand, people choose where to live and firms choose where to locate, making density endogenous to local conditions. Despite some notable contributions, researchers are still searching for compelling evidence that establishes a link between density and information diffusion (Glaeser and Gottlieb, 2009).

This paper tackles both challenges. In response to the endogeneity challenge, I exploit a unique national policy in China that requires all residential buildings to receive two hours of sunshine a day in winter. This policy creates varying degrees of restriction on the density of buildings across latitudes, as northern buildings have to be further apart from each other due to longer shadows at higher latitudes. The resulting differences in building density exogenously induce differences in population density across latitudes. China thus provides an ideal setting to study the implications of urban density, due to

the presence of this source of variation and the country's expansive territory.

To measure information diffusion, this paper uses observed dynamics of user activity on Local Leader Message Board, a national online petition platform, where a separate board exists for each city, and all postings and government replies are publicly viewable. I study how the volume of postings on a given topic responds to local government's reply rate to the same topic. The advantage of this setting is that I observe the same information shocks as the users under study, which allows me to trace out gradual increases in postings over time consistent with diffusion of information about local government's shifting priorities.

To understand the implications of the sunlight policy, I start by characterizing the behavior of developers in choosing building height given the policy. Under the assumption that it is increasingly more costly to build an extra floor, profit-maximizing developers will choose to lay out buildings in such a fashion that allows just enough sunshine to reach each building. The developer's optimal building height is then determined by a balance between increased sales from more floor space and the corresponding increases in construction costs. The increase in floor space associated with a unit increase in height is connected to the angle of sunlight, and turns out to be higher in the south. Developers in the south are thus predicted to choose to build higher buildings, all else being equal.

This prediction is tested with 2005 1% census data, which record whether a household lives in a high-rise (with at least seven floors) and the construction year of the building. Before the policy was implemented, there is no systematic difference in the prevalence of high-rises across latitudes. The prevalence of high-rises starts to diverge in 1993, exactly the year when the sunlight policy was enacted. Households living in buildings constructed after 1993 are more likely to be in a high-rise in the south than in the north, after a host of residence, household and city characteristics are controlled for. This result is robust to the inclusion of province-specific trends. That is, within the same province, southern cities constructed more high-rises than northern cities after 1993. Moreover,

southern and northern cities did not experience systematically different growth rates in urban population, employment or GDP, alleviating the concern that differential economic growth may be driving the differential growth of high-rises.

The diverging prevalence of high-rises suggests that the policy was enforced. Compared to building height, we are more interested in how much floor space is allowed to be built, which determines housing supply. The density of floor space can be succinctly summarized in the value of Floor Area Ratio (FAR), defined as the amount of floor space built on one unit area of land. Since developers' optimal FAR is linked to the solar elevation angle, the optimal FAR level varies across latitudes.

The Chinese government uses FAR upper limits as the primary tool to restrict urban density (Brueckner et al., 2017), as is also the case in India (Bertaud and Brueckner, 2004; Harari, 2017), Japan (Asami et al., 2006), and many other countries around the world (Brueckner and Lall, 2015). If the government seeks to maximize revenues from land sales, as could be a reasonable first-order approximation in the Chinese context (Chen and Kung, 2016), it will set regulatory FAR at developers' optimal level. A comprehensive dataset of land transactions covering a majority of new housing development in urban China during 2007-2016 and containing FAR upper limits for individual projects reveals that the cross-latitude variation in regulatory FAR can be described quite well by a formula implied by developers' profit-maximizing behavior. Including a rich set of controls at project and city levels does little to change the correlation between observed FAR and that predicted by the model.

The strong correlation between observed and predicted FARs suggests that the sunlight requirement is often the binding contraint on density, surpassing other considerations such as ventilation, views, parking space, fire safety and households' demand for open space. To see whether the sunlight requirement is *always* a binding constraint on density, I explore deviations from the policy, and find that the regulatory FAR of housing projects on larger plots and in richer cities has a significantly weaker correlation with

that predicted by the sunlight policy. These projects are subject to stricter development restrictions (i.e., lower FARs) than required by the sunlight policy, possibly reflecting higher demand for open space by richer households and better transportation infrastructure in richer cities. All in all, my estimates suggest that moving across the interquartile range of latitudes, regulatory FAR would increase by 15%, all else being equal.

Under these exogenously set FAR restrictions, similarly dense 1-km^2 grid cells in 1990 have diverged in terms of population density by 2010, with southern cells becoming significantly denser. This holds within provinces and when controlling for initial city characteristics. A standard urban economics model with homogeneous households predicts higher land prices where population density is higher. Consistent with this prediction, I find that among grid cells with similar population densities in 1990 and similar nighttime light luminosities in 1992, those in the south had higher residential land prices during 2007-2016, the sample period of my land transactions data. Cells in the south also had higher commercial land prices, consistent with a model featuring an integrated land market where land is allocated to its most profitable use. In a world where the residential and commercial land markets are separate due to zoning restrictions, this is consistent with greater market access to consumers in denser places being capitalized into commercial land rents, which has been modeled in Ahlfeldt et al. (2015) and Donaldson and Hornbeck (2016), among others. Interestingly, I find that commercial FAR is also higher in the south, even though the sunlight policy does not apply to commercial development. This can be explained by an elastic supply of commercial floor space in response to higher rents.

The exogenous variation in density driven by the sunlight policy is then used to examine the relationship between density and information transmission. I examine user behavior dynamics on Local Leader Message Board, exploiting differential trajectories of posts on different topics (issues) discovered with an LDA model (Blei et al., 2003). There is clear evidence of responses to shifts in government attention. After the local government

increases their reply rate to a certain topic, volume of posts from the city on the same topic gradually increases, exhibiting an S shape over time, which is suggestive of a local diffusion of information about the government's shifting priorities, where people learn these shifts from the replies and increase their postings due to either strategic considerations or psychological encouragements. Consistent with this interpretation, I find that grievance-related topics are significantly more responsive to changes in government reply than non-grievance topics. The observed responses are not caused by a reverse causality where the local government replies more to more frequently raised issues, as increases in postings do not anticipate increases in reply rate.

Consistent with density contributing to more rapid information transmission, posts in southern cities respond systematically faster to changes in government reply than in northern cities. This is true when a host of city-level economic, education and telecommunication infrastructure characteristics are controlled for, and when only using differential responses within provinces. In terms of magnitude, estimates suggest that in the first two quarters after a reply rate increase, a city at the 25th percentile of latitude would experience a cumulative increase in postings that is 2.7 times as large as a city at the 75th percentile of latitude. Survey data further provide direct evidence that similar individuals with similar levels of interest in public issues are more likely to gossip and hear gossip about such issues in the south than in the north, lending support to the conjecture that density is conducive to the spread of word of mouth.

To address alternative explanations of the observed patterns, I present evidence that citizens do not differ systematically across latitudes in their attitudes towards the government, as measured by their beliefs about the role of government in various domains and trust in various government institutions, addressing the concern that different tendencies to use the platform may be driving differential responses across latitudes. There is also no evidence of systematic cross-latitude differences in internet using behavior, alleviating the concern that information may be transmitted online instead of offline.

This paper contributes to the literature on the relationship between density and knowledge spillovers, which has been studied both in the context of agglomeration economies (Rosenthal and Strange, 2004) and human capital externalities (Moretti, 2004). Density is an endogenous outcome for which reasonably exogenous sources of variation are not easy to find. Exceptions are Combes et al. (2010), who use soil quality to instrument for historical settlement patterns, and Rosenthal and Strange (2008), who use geology (landslide hazard, seismic hazard, and the presence of sedimentary rock) to instrument for the prevalence of high buildings. I introduce a new instrument that is shown to have a robust first stage. To my knowledge, Bertaud (2018) and Zhang (2017) are the only works in economics that have discussed the sunlight policy in China, and neither of them uses detailed project-level data to verify that the policy is indeed currently enforced and appears to be a binding constraint on urban density.

The transmission of information and knowledge is hard to observe, prompting scholars to rely on implications of knowledge spillovers including variation in wages and rents across cities with different characteristics (Rauch, 1993; Glaeser and Mare, 2001) or the location choice of firms (Audretsch and Feldman, 1996; Arzaghi and Henderson, 2008) instead of directly measuring them. An important contribution is Jaffe et al. (1993), who study the paper trail of patent citations to identify knowledge spillovers. I measure information diffusion using postings on a public platform, which has the benefit of being directly observable. While this specific type of information is not directly related to productive activities, it has potential implications for government accountability and the quality of governance.[1]

This paper also contributes to the literature on the consequences of residential development regulations (Gyourko and Molloy, 2014; Turner et al., 2014). While development

[1]Campante and Do (2014) shows that in US states where the capital city is situated in a sparsely-populated area, newspapers cover state politics less, voters have less interest and knowledge about state politics, and the level of corrpution is higher. In an international context, Campante et al. (2014) shows that isolated capitals are associated with misgovernance in less democratic countries.

restrictions are usually complicated and varied, necessitating use of indirect measures such as the gap between housing price and marginal construction cost (Glaeser et al., 2005), FAR upper limits readily quantify the degree of restriction on floor space construction. Unlike in the US where zoning ordinances are set at local level, influenced by the interest of local residents and therefore more likely to be plagued by endogeneity, the sunlight policy in China is a national policy that exogenously shifts urban density across a country with the world's largest urban population. In addition to its broad impacts, the centralized nature of the system makes available a dataset that covers a majority of new urban construction projects across the country during the last decade. Drawing on this and other sources, this paper shows that zoning regulations can have measurable impacts on urban density and the speed of information diffusion.

The chapter proceeds as follows. Section 1.2 discusses the motivation behind the sunlight policy, how the policy is enforced, as well as some background on the urban land market. Section 3.2 describes the main datasets used in this chapter. Section 1.4 derives a relationship between latitude and optimal FAR and empirically looks at the effect of the sunlight policy on building height. The section also shows that regulatory FAR across China can be characterized quite well by a formula linking FAR to latitude through the solar elevation angle. Section 1.5 shows the dynamic responses of postings to changes in government reply rate, and reports the systematic differences in response speed across latitudes. Section 1.6 discusses alternative explanations that could give rise to the observed patterns. Finally, Section 1.7 concludes.

1.2 Institutional Background

1.2.1 The Sunlight Policy: Origins and Evolution

Sunshine is an important factor in the culture of *feng shui*,[2] and is valuable to Chinese households for practical purposes including drying clothes[3] and keeping homes warm in winter.[4] Soviet building practices[5] could also have influenced Chinese policymakers' perception about the importance of citizens' right to sunshine. However, there was no law governing building distances in relation to sunlight prior to the 1990s, partly because the pre-1990s urban population was relatively small and land value was low, making sunlight easily guaranteed for residents.

With the unleashing of the commercial real estate market after 1998 and the beginning of large-scale housing construction (Wang, 2011; Fang et al., 2016), the sunlight right of residents started to be taken more seriously. *Urban Residential Planning and Design Ordinance* (GB50180-93) issued by the Department of Construction in 1993 gave this policy a legal status. It requires that the lowest level of any residential building receive at least two hours of sunshine during a specific winter day on the Chinese lunisolar calendar.[6] Having undergone a few revisions, this ordinance is still effective today, wherein the sunlight requirement has not changed since 1993.

---

[2] A traditional system of beliefs and practice regarding residence location and layout.

[3] Dryers are a very rare sight in contemporary China. A survey (Brockett et al., 2002) conducted in 1999 to 251 households in five Chinese cities across climate zones found that only two households owned a dryer.

[4] Centralized heating is only available to the northern part of the country (Almond et al., 2009).

[5] According to a report by a US delegation to the USSR (Wright, 1971), "Two prime planning concerns for the Soviets are light (three hours of sunlight per room at the March solstice), and breathing room."

[6] This day is Major Cold for most cities, which is the day when the sun is at 300 degrees ecliptic longitude and falls between January 19 and January 21 every year. There are some alternative specifications depending on the climate zone the building is in and whether the building is in a "large" city or a "small or median" city. Buildings in "small or median" cities have to receive three hours of sunshine on Major Cold, unless they are in climate zones IV, V or VI, which correspond roughly to the very south of China and the Tibetan Plateau, in which case they have to receive at least one hour of sunshine on the day of Winter Solstice.

Local building code issued by provinces and prefectures respect and reflect this national policy. These documents often contain clauses governing the distance between buildings in relation to sunlight, and in many cases explicitly stipulate a "sunlight distance coefficient", referring to the required ratio of distance between two adjacent buildings to building height. I collect these coefficients from over 200 prefectural building code, and plot them against the latitude of the prefectures in Figure 1.1.[7] There is a strong positive correlation between the sunlight distance coefficient and latitude - buildings in northern cities are required to be further away from each other. The variation across cities is large. Buildings in Changchun (43.79°N) are required to be 1.95 times as far apart as buildings in Kunming (25.19°N).

Theoretically, the minimum sunlight distance coefficient that guarantees sunlight for the first floor of each building should be equal to the cotangent of solar elevation angle at a location, which depends on the latitude of that location.[8] This theoretical sunlight distance coefficient curve (equal to the cotangent of solar elevation angle) is overlaid in Figure 1.1, and achieves an R^2 of 0.99 with the coefficients specified in actual local regulations.[9]

---

[7]Specifically, shown in this figure are sunlight distance coefficients in different cities for residential buildings below seven stories in new urban areas. Buildings above seven stories are regulated via minimum distances (in meters) between buildings. There is typically a separate sunlight distance coefficient for old city centers where the sunlight requirement is slightly relaxed, the spatial extent of which varies from city to city. In the sample, the mean ratio of central city sunlight distance coefficient to new urban area sunlight distance coefficient is 0.86.

[8]Let ϕ denote the latitude (in radians). The formula linking solar elevation angle to latitude is

$$\sin \alpha = \sin \phi \sin \delta + \cos \phi \cos \delta \cos h$$

where α is the solar elevation angle. δ is the declination of the sun, which on the day of Major Cold is roughly -20 degrees ($-\frac{\pi}{9}$). h is the hour angle. In order for a building to receive at least two hours of sunshine on that day, sunlight must reach the building 11 am at the latest. We therefore evaluate α at 11 am and set $h = -\frac{\pi}{12}$ (-15 degrees). Therefore,

$$\alpha(\phi) = \arcsin(\sin(-\frac{\pi}{9}) \sin \phi + \cos(-\frac{\pi}{9}) \cos(-\frac{\pi}{12}) \cos \phi)$$

[9]It is worth noting that cities in southernmost and northernmost China seem to follow the national rule less closely than the rest of the country, at least according to these coefficients specified for buildings below seven stories. Cities in northeast China (north of 40°N) set local building distance coefficients

Are these local regulations followed in practice? Having a construction plan that satisfies the sunlight requirement for each building is a prerequisite for obtaining a building permit. Enforcement of the rule is sophisticated and involves specialized computer software. In many cities, developers are required to use an officially-sanctioned sunlight analysis software to model the changing sunlit conditions of buildings throughout the day, and regulators use this software to evaluate hours of sunlight exposure building by building.[10]

The fact that citizens can take their sunlight-blocking neighbor to court further increases the enforceability of the sunlight policy. The right to sunlight was enshrined in the Property Law enacted in 2007. Since 2007, there have been at least 2,000 court cases across China regarding sunlight rights.[11][12]

In Section 1.4.3, I present evidence from the regulatory floor area ratio specified for a large fraction of all urban housing projects constructed across China between 2007 and 2016, which supports the notion that these local sunlight requirements are followed quite closely in setting regulatory floor area ratio for individual housing projects. Floor area ratio is defined as the ratio of total floor space on a piece of land to the area of that piece. FAR will be high when buildings are either high or take up a large fraction of land on which they sit. As such, FAR directly measures building density.

---

systematically lower than that required by the national sunlight rule, due to the high cost of providing utilities and infrastructure for buildings too far apart, which is particularly challenging as northeastern municipal governments generally face varing degrees of financial distress (See Letter Regarding Compliance of Urban Residence Sunlight Standards in Heilongjiang, an official document issued by the national managing committee of the sunlight policy after their field trip to the northeast in 2007). On the other hand, cities to the south of 25°N typically set their sunlight distance coefficient at 1.0, even though in theory they can further decrease the distance without violating the sunlight right of residents, presumably because other considerations in determining building distances (ventilation, views, green space, parking space, fire safety, etc.) start to kick in when buildings are close enough to each other.

[10]Sunlight Parameter Standards for Buildings (GB/T 50947-2014), Ministry of Housing and Urban-Rural Development, PRC, 2014.

[11]I search for court cases whose cause is "neighboring sunlight dispute" on three of China's leading court case disclosure platforms, and find over 2,000 cases on each of the three platforms. The three platforms are *openlaw.cn*, *China Judgements Online* and *itslaw.com*.

[12]In a headline news story, a man won a case against Zifeng Tower, the tallest skyscraper in Nanjing, for blocking sunlight to his apartment.

1.2.2 The Urban Land Market

Residential construction in urban China is predominantly channeled through real estate developers instead of individual home buyers. Local governments transfer large land lots to developers, who in turn build apartments that they sell to home buyers.[13] My dataset sheds light on this phenomenon. The median size of a greenfield residential lot in the sample is 1.46 hectares (157,153 square feet), which is much larger than the lot needed for building a typical single-family residence. This number is 3.21 hectares in the 314 prefectures, which are the larger cities. The 20th percentile of the size of a greenfield residential lot is 0.07 hectares (7500 square feet). If we think of 0.07 hectares as about the right size for building a single-family residence, lots larger than 0.07 hectares account for 99.9% of total area transferred. Because the lots are transferred to developers instead of ordinary home buyers, they face financial incentives to build up to the legally allowed FAR upper limit. This justifies studying the regulatory FAR upper limit as a key determinant of urban density in China.

Urban land in China is state-owned. During urban expansion, local governments expropriate rural land surrounding cities from farmers, at which moment they obtain ownership over that land. They then transfer the use right of land to developers or industrial firms for commercial, industrial and residential development. The tenure length of this use right is generally fixed and varies by use type (40 years for commercial, 50 years for industrial, 70 years for residential).

There are three legal procedures in which local governments transfer land to users: allotment, auction and contract. Allotment (22.6% of transactions in the sample) applies to land for public infrastructure and buildings, where the lots are transferred at close to zero cost.[14]

[13] Fang et al. (2016) uses the fact that most housing construction is organized as large residential communities to propose a housing price index that measures time variation of price within communities.

[14] 66.5% of allotted lots in the sample involve zero payment, excluding observations where the price is missing, which is more likely than not to be missing because it is zero.

Auction is the legally required means for transferring commercial and residential land. Under the most prevalent auction form (*guapai*), the lot for transfer is advertised beforehand, and individuals or firms who want to bid for the lot post their bids online after passing local government's qualification checks and making a deposit. The auction generally lasts ten days and happens online, during which period bidders can continually submit new bids, and the highest bid is updated online. At the end of the bidding period if there are two or more bidders who offer the same price, an on-site auction is held to determine the winner.

Finally, contract mainly applies to transactions involving changes of land status. The Appendices contain a detailed discussion of situations where this is applicable.

1.3 Description of Data

To increase transparency in land transactions, the Ministry of Land and Resources[15] requires local governments to report all land transactions, which are subsequently disclosed on the Ministry's website.[16] Disclosed information includes location, zoned use, price, size, regulatory FAR upper limit (and in some of the industrial land transfers lower limit), identity of receiving party, means of transfer, land grade,[17] date of contract, etc. All the transactions are publicly viewable. I scraped all records from the website in early 2017. The resulting dataset contains over 1.3 million transactions from across China. Since all construction on greenfield urban land has to obtain land from the government and therefore is registered in the dataset, I observe close to the universe of new urban construction in the 2007-2016 period.[18]

[15] In 2018 the Ministry was consolidated into a new Ministry of Natural Resources.

[16] http://landchina.mlr.gov.cn.

[17] Government-accessed land quality, which reflects, among other things, whether the lot is connected to road and telecommunication networks and serviced by utilities including heating, electricity, water, gas and sewage.

[18] I cross-checked total area of land transfered in my dataset with official numbers published in Land and Resources Yearbooks. My dataset covers 69.8% in area of all land transfered during 2007-2015. Land

Of the land parcels transfered in or after 2007, around 40% of them are zoned for residential uses. These residential parcels have a median price of 318 RMB per square meter and a median FAR upper limit of 2.29. Using Baidu Map API, I geocoded all parcels, translating their addresses into longitudes and latitudes. 98.4% of parcels were successfully located.

Another source of data I utilize is the 2005 1% population census administered by the National Bureau of Statistics. The census uses a three-stage cluster sampling method to ensure representativeness (Cui et al., 2005) and contains rich information at both household and individual levels. Important for my purposes, the census contains information on whether a household lives in a high-rise (buildings with at least seven floors) and construction year of the building.

To examine information diffusion, I make use of postings on the Local Leader Message Board,[19] a national online platform where people register their complaints and suggestions to government officials in their local area. The platform is managed by People's Daily, China's central official medium, and contains sub-boards for all administrative units (provinces, prefectures and counties). Local governments can respond to messages posted to their board, but ordinary users cannot respond to each other's postings. All the postings and government replies for all localities are openly viewable.[20] In an earlier project (Jiang et al., Forthcoming), we scraped the contents of all postings and replies up until early 2016 along with auxiliary information including time of posting and locality posted to.[21] The resulting dataset contains around 900,000 postings.

---

transactions prior to 2007 are not covered well by the Ministry of Land and Resources website. Therefore I focus my analysis on transactions that happened in or after 2007 in this paper.

[19] http://liuyan.people.com.cn.

[20] Interviews with LLMB staff suggest that censorship on the platform is much less severe than for social media, partly because the platform is managed by a central organization who has little incentive to cover up for local officials, and indeed may have an incentive to elicit information about local situations.

[21] We matched messages posted to provincial boards to cities by finding city names in the text. In cases where multiple cities are mentioned, the most frequently mentioned city is chosen. Only 8.7% of posts cannot be matched to a city.

Other sources of data, including 1-km^2-resolution population density, nighttime lights, terrain ruggedness, and climatic conditions, are summarized in Table A1.

1.4 Effects of the Policy on Building Height and FAR

Developers want to maximize their profits under the sunlight constraint. In order to guarantee sunlight to the next building, taller buildings have to be accompanied by larger distances between them. Even so, this section shows that increasing building height will increase total floor space available under the sunlight constraint. The marginal benefit of increasing building height will be larger in the south than the north, all else equal. Consequently, developers in the south will optimally choose to build taller than developers in the north. I then test this hypothesis using 2005 census data recording whether households live in a high-rise (with at least seven floors) and construction year of the buildings.

We are ultimately more interested in residential FAR than building height, as FAR determines the amount of residential space available. The analysis of developer behavior also implies that residential FAR will be higher in the south than in the north, and yields a formula that links latitude to residential FAR. Taken to housing projects constructed across China 2007-2016, this formula proves to describe the cross-latitude variation in regulatory FAR set for individual housing projects quite well.

1.4.1 Developer Behavior under the Sunlight Policy

Developers operate in perfectly competitive markets. At location i, they take housing price there p_i^H as given. Revenue from constructing floor space F_i^r will thus be $p_i^H F_i^r$. Suppose lot size is fixed by the government.

Construction cost $CC(F_i^r, h_i)$ depends on total floor space constructed F_i^r and the height of buildings h_i, and satisfies $\frac{\partial CC}{\partial F} > 0, \frac{\partial CC}{\partial h} > 0$. It is increasingly more costly to make buildings taller, so that a low building hosting the same floor space as a tall building will be less costly to build.

The developer's profit maximization problem can be considered in two steps. She chooses a layout of buildings that minimizes construction costs to achieve a given desired FAR (i.e., total floor space), and she then chooses a profit-maximizing FAR.

The configuration that minimizes construction costs to achieve a given FAR under the sunlight policy is to place buildings in ranks and files reminiscent of Le Corbusier's *Radiant City* plan such that there is just enough room for sunlight to reach the first floor of each building.[22] Figure 1.3 illustrates this configuration. Leaving more space between buildings will have to be compensated with taller buildings to achieve the same amount of floor space, which is more costly as the marginal cost of height is increasing.

We calulate FAR associated with the layout on the left of Figure 1.5 where the length of a building takes up one full dimension of the lot. Although it may seem odd to construct such elongated buildings, note that a more realistic staggered layout such as that on the right of Figure 1.5 actually achieves the same FAR.

The relationship between distance d and height h is given by solar elevation angle α,

$$h = d \tan\alpha$$

Let l be the length of a parcel and l_0 be the width of a building. l_0 is fixed due to housing unit layout considerations. Then the number of buildings that can be constructed on the parcel is $N = \frac{l}{l_0+d}$. Floor area ratio is equal to the fraction of land taken up by buildings multiplied by number of floors. Let h_f be the height of a floor, which is also

[22] We ignore idiosyncrasies in the shape of buildings, e.g., setbacks.

fixed. We have

$$\text{FAR} = \frac{Nl_0}{l}\frac{h}{h_f} = \frac{l_0}{h_f}\frac{h}{l_0 + \frac{h}{\tan\alpha}} = \frac{l_0}{h_f}\frac{1}{\frac{l_0}{h} + \frac{1}{\tan\alpha}}$$

.

We see that under this configuration of buildings, it is possible to increase floor space by increasing building height while adhering to the sunlight policy. Building taller is however more expensive. Let S be lot size. The marginal benefit of increasing height is given by

$$\text{MB}_h = p^H S\frac{d\text{FAR}}{dh} = \frac{p^H S}{h_f}(\frac{l_0}{l_0 + \frac{h}{\tan\alpha}})^2$$

We see that the marginal benefit of height is decreasing in h, i.e., total revenue as a function of height is concave. Moreover, the marginal benefit of height is increasing in solar elevation angle α, which implies larger marginal benefits of height in the south.

The marginal cost of height is given by

$$\text{MC}_h = \frac{dCC(F^r, h)}{dh} = \frac{\partial CC}{\partial F^r}\frac{dF^r}{dh} + \frac{\partial CC}{\partial h} = \frac{\partial CC}{\partial F^r}S\frac{d\text{FAR}}{dh} + \frac{\partial CC}{\partial h}$$

Suppose that holding height fixed, the marginal cost of building one extra square foot is constant, i.e., $\frac{\partial CC}{\partial F^r} = \eta$. The developer's optimal building height under the sunlight policy is a level h^* such that $\text{MB}_h = \text{MC}_h$, i.e.,

$$\frac{(p^H - \eta)S}{h_f}(\frac{l_0}{l_0 + \frac{h^*}{\tan\alpha}})^2 = \frac{\partial CC}{\partial h} \tag{1.1}$$

From this, we see that given the sunlight policy, developers in the south (under larger α) will choose to build taller than their counterparts in the north, under the same construction cost schedule determined by local wage levels and construction technology and the same housing price.

1.4.2 Evidence from Building Height

In this subsection, I test the hypothesis that developers in the south will build taller under the sunlight policy *ceteris paribus* using census data from 2005, which record whether a household lives in a high-rise (with at least seven floors) and construction year of the building. Specifically, I estimate the following equation:

$$\text{High-Rise}_{hct} = \lambda_{\text{prov}(c)} + \gamma_1 \text{Lat}_c + \gamma_2 \text{Post}_t + \gamma_3 \text{Lat}_c \times \text{Post}_t + \beta_1' X_c + \beta_2' X_h + \epsilon_{hct}$$

High-Rise_{hct} is an indicator for whether household h in city c lives in a high-rise conditional on the building being constructed in year t. $\lambda_{\text{prov}(c)}$ are province fixed effects, which control for province-specific determinants of building height. Post_t is an indicator for whether the building construction year is no earlier than 1993, the year when the sunlight policy was written into law. X_c is a vector of city-level controls in 2005 that captures determinants of housing prices and local wages,[23] including urban population, population density, employment, GDP per capita and share of land with above 15% slopes within a 10-km radius from city center.[24] Cities surrounded by rugged terrain tend to have a less elastic housing supply and higher housing prices (Saiz, 2010).

Ideally we would like to observe all buildings built in all years, but the nature of population census data determines that we can only observe occupied buildings. To address the concern that households in the north and the south sort into buildings of different vintages differentially, the equation includes a vector of household-level controls X_h. These include both residence characteristics and demographics. Residence controls contain use porpose of residence (just for living or for both living and production), log number of rooms, log square footage, whether shared with other households, access to tap water, presence of kitchen, toilet and bath facilities, type of fuel and how the residence

[23] Direct data on housing prices and wages are not available for counties.

[24] City center is defined as the brightest 1-km cell in terms of 2005 nighttime light. In cases where multiple cells are equally bright, their geometric center is calculated.

was obtained (rent or bought, SOE housing, affordable housing or commercial housing). Demographics contain log number of household members and the following characteristics of household head: gender, age, ethnicity, *hukou* registration place, type of *hukou* (rural or urban), literacy, education, whether worked for more than one hour last week, log income, source of income and marital status.

In other words, we test whether similar households residing in similar residences in similar cities have different probabilities of living in a high-rise depending on latitude of the city, and whether that difference only occurs for buildings built after the sunlight policy was put in place. Table 1.1 shows regression estimates. Column 1 is a parsimonious specification containing just province fixed effects. We see that households living in buildings constructed after 1993 are over 30% more likely to be in a high-rise, reflecting China's urbanization process. The coefficient on latitude is not significantly different from 0, suggesting that the prevalence of high-rises was not different across latitudes before the sunlight policy. However, the negative coefficient on the interaction of post and latitude suggests that buildings constructed after the policy are more likely to be high-rises in the south than in the north. Magnitude of the coefficient indicates that across the interquartile range of latitudes (moving from 39.2 degrees to 25.4 degrees), households would become 5.5% more likely to live in a high-rise conditional on the building being constructed after 1993.

Column 2 adds 2005 city-level controls. Coefficient on post decreases slightly, suggesting that part of the increase is driven by changes in city composition within different building vintages. However, the interaction remains negative and significant. It is worth noting that surrounding slopes significantly increase building height, as expected from the model where slopes increase housing prices and therefore the marginal benefit of height. Columns 3 and 4 add residence controls and household controls consecutively. The coefficients remain quite stable. Column 5 further includes city fixed effects, which absorb latitude and city-level controls. We see that the coefficients of interest remain stable.

Column 6 includes both province fixed effects and their interactions with the post dummy. This allows for province-specific trends in building height, and identifies the effect of policy solely from within-province differential growth across latitudes. This limited variation induces much less precise estimates. Reassuringly, the point estimates remain almost unchanged. This indicates that even within the same province, a southern city exhibits faster increase of high-rises than a northern city after the policy was in place.

Is the differential growth in building height across latitudes caused by differential economic growth? City-level economic outcomes are not available before 1994, which prevents me from directly including them in the regression. However, we can test whether they exhibit differential growth in the period we do have data for. I estimate the following equation:

$$y_{ct} = \lambda_c + \delta_t + \gamma_t \times \text{Latitude}_c + \epsilon_{ct}$$

where y_{ct} is an economic outcome (log urban population, log employment, log GDP or log GDP per capita) for city c in year t. γ_t captures any differential economic growth across latitudes. Each regression is run on a panel balanced on the outcome, i.e., cities with any missing values in the outcome are excluded from the regression.

Table 1.2 shows the estimated set of γ_t. From columns 1 to 4, we see that there is no evidence of differential growth across latitudes in urban population, employment or GDP, although GDP per capita does seem to grow faster in the north between 2002 and 2005. Columns 5 to 8 include province by year fixed effects to detect differential economic growth within provinces. There is no evidence for differential growth in any of the four variables within provinces. Taken together, these findings suggest that the differential changes in the prevalence of high-rise buildings are likely due to differential incentives created by the sunlight policy.

Could this differential change in building height have been caused by other policy events than the sunlight policy? Afterall, mid- to late-90s was also the time when the government privatized housing and kick-started the real estate sector. Fang et al. (2016)

indicates that housing construction only started to grow rapidly after 1998, when the central government completely abolished the traditional model of providing housing as an in-kind transfer to SOE employees, and started to provide subsidized residential mortgages. I examine the timing of building height change in the data to see whether the change appears to have happened in 1998 or in 1993 when the sunlight policy was introduced. To be specific, I estimate

$$\text{High-Rise}_{hct} = \lambda_c + \delta_t + \gamma_t \text{Latitude}_c + \beta_1' X_c + \beta_2' X_h + \epsilon_{hct}$$

Figure 1.7 plots the trajectory of $\delta_t + \gamma_t \text{Latitude}_c$ for a northern city (75th percentile of household latitude) and a southern city (25th percentile of household latitude). We see that the two trends started to diverge almost exactly when the sunlight policy was introduced and years before 1998. This lends support to the claim that the sunlight policy drove a differential growth of high-rises in the south.

1.4.3 Evidence from FAR

So far we have seen the effects of the sunlight policy on building height. We are more interested in the floor area ratio of buildings, as that determines how much floor space is available for households to consume. Panel A of Figure 1.9 shows a binned scatterplot of log regulatory floor area ratio of individual residential projects over latitude. There appears to be a strong negative relationship between FAR and latitude - lots in the south are allowed to accommodate more floor space. Even though no controls are included in this exercise, the binned conditional means fall around the fitted negative trend quite closely.[25]

[25]The Appendix contains scatterplots of regulatory FAR over latitude using the raw data of individual lots along with their non-parametric fits (Figure A5), where we can have an idea of the variation at lot level. The figure also breaks down data into eastern, central and western China. In all regions the negative relationship between regulatory FAR and latitude is present.

Recall that developer's optimal FAR at a given latitude is given by

$$\text{FAR}^* = \frac{l_0}{h_f} \frac{1}{\frac{l_0}{h^*} + \frac{1}{\tan\alpha}}$$

From the land transactions we observe regulatory FAR upper limit set for individual housing projects. The government transfers use right of land parcels through auctions. If the government's goal is to maximize their auction revenue, they will set regulatory FAR at exactly developers' desired level under the sunlight policy, as that will maximize developers' bids. In this subsection I evaluate whether this expression describes the cross-latitude variation in regulatory FAR well.

By Equation 1.1, the optimal height h^* depends on the construction cost schedule and housing prices, which we do not have sufficient information about. Therefore in this section I ignore potential developer optimization with respect to building height, and make the approximating assumption that all buildings are of the same size. Width of a building is assumed to be 14 meters, and height is assumed to be 21 meters (around 7 floors).[26]

This approximation only works well if the determinants of optimal building height do not differ systematically across latitudes. To ensure this, I include flexible geographical fixed effects and a rich set of controls in the following estimating equation. The identifying assumption is that within relatively small spatial cells, latitude is not correlated with unobserved determinants of floor area ratio conditional on a rich set of controls.

$$\log \text{FAR}_{pct} = -\beta \log\left(\frac{2}{3} + \frac{1}{\tan\alpha_p}\right) + \lambda_{g(\text{lng}_p, \text{lat}_p)} + \gamma_1' X_p + \gamma_2' X_c + \gamma_3' X_t + \epsilon_{pgc} \qquad (1.2)$$

If local governments set floor area ratios to the maximum level possible that just suffices to guarantee every resident's right to sunlight, we would expect β to be 1. The extent to which β is lower than 1 measures the extent to which local governments deviate

[26] These approximations are provided by a Chinese architect.

from the sunlight policy. Deviations can take two forms - they can either set an FAR that is lower than that required by the sunlight policy,[27] or they can allow developers to build more floor space than allowed under the sunlight policy. Only the latter is a violation of the national policy, but both would cause the correlation between FAR and our formula to be lower than one.

Turning to the controls in Equation 1.2, I include fine-grained longitude-latitude cell fixed effects to control for unobservables. In the main specifications, I use 0.1 degrees longitude by 2 degrees latitude fixed effects, as we want to maintain some variation in latitude within these cells. $\lambda_{g(\text{lng}_p,\text{lat}_p)}$ denote these longitude-latitude cell fixed effects.

X_p is a vector of project-level controls, including whether the lot is greenfield (new land) or brown-field, type of residential building (fixed effects for luxury villas, regular housing, median and low price housing, economy housing,[28] affordable rental housing,[29] or public rental housing[30]), whether the lot is transferred through allotment, contract or auction, fixed effects for land grade, log size of lot, log distance to 2005 city center,[31] and 2005 nighttime lights. If these covariates vary systematically across latitudes and influence either developers' demand for FAR or the government's desired FAR level, omitting them

[27]There are many other potential considerations in setting a regulatory FAR upper limit than complying with the sunlight rule (ventilation, views, green space, unit layout, fire safety, etc.). In practice, the sunlight rule may not be binding in sparsely-populated places. The *Urban planning and management technical specifications* of Gansu province exemplifies this well, a paragraph in which reads: "In the enforcement of this regulation, when facing projects of different scales and localities of different characteristics, the relevant targets are not unique, but should be determined within a range in accordance with other requirements in this document... The vast Gobi and oasis areas to the west of Yellow River are suitable for low-density construction, while middle and eastern parts of our province require construction of relatively high density."

[28]*jingji shiyong fang*. These are government-subsidized housing, where the government grants land to developers at very low prices (the median price of such land is 14.29 yuan per square meter in the sample) and waives related municipal fees, in exchange for their charging a lower price to home buyers.

[29]*lianzu fang*, targeted at households eligible for urban minimum living standards assistance (*dibao*) and facing difficulty with housing.

[30]*gongzu fang*, lottery-based rental housing targeted at households of relatively low income but ineligible for affordable rental housing, migrant workers and people who are newly employed.

[31]City center is again defined as the brightest 1-km cell in terms of 2005 nighttime light. In cases where multiple cells are equally bright, their geometric center is used.

will induce bias in the estimation of β.

X_c is a vector of covariates at the city level in 2006[32], a city being defined as either a prefecture jurisdiction area[33] or a county. The controls include urban population, GDP per capita, population density, employment, share of land with slopes above 15% within a 10-km radius from city center, climatic controls including mean temperature in January, mean temperature in July, and average annual precipitation over the period of 1970-2000.

Finally, X_t is a vector of time covariates including year of transaction and month of year fixed effects. Year fixed effects control for both business cycles in real estate development and secular trends, while month of year fixed effects control for possible seasonality in the demand for floor space.

Before turning to regression estimates, panel B of Figure 1.9 shows a binned scatterplot where both log FAR and latitude are residualized within these longitude-latitude cells and over the aforementioned controls. Residualizing within the fine cells reduces the policy-induced variation compared to noise, and therefore leads to more dispersion in the data. However, we still see a clear negative relationship between latitude and FAR.

Table 1.3 presents the regression results, where I call $-\log\left(\frac{2}{3}+\frac{1}{\tan\alpha_p}\right)$ "predicted log FAR". In column 1 where log regulatory FAR is simply regressed on predicted log FAR with no controls, β is estimated at 0.68, and is highly significant. Unsurprisingly, β is estimated to be less than 1, but the magnitude of 0.68 suggests that the model does a reasonably good job at characterizing mean FAR conditional on latitude. The r-squared of this regression is just 0.055, due to dispersion in FAR at a given latitude. Through the lens of the model, this can be rationalized with variation in optimal height h^* due to different demand conditions and construction costs at different locations within a latitude.

[32]I choose to control for these city level characteristics in 2006 instead of later years to make these covariates more closely resemble predetermined variables that ensure consistent estimation, since I use land transactions in or after 2007 in the regressions.

[33]*shixiaqu*, including districts governed by a prefecture, but excluding counties under the prefecture. This is an appropriate unit of analysis, as a prefecture directly controls only land in its districts, while counties have autonomy in managing their land. The prefecture jurisdiction area roughly corresponds to the urban core of a prefecture.

Adding longitude-latitude-cell fixed effects, time fixed effects, lot-specific controls and city economic controls sequentially in columns 2 through 4 changes the estimated magnitude of β only slightly. It is interesting to note that larger lots tend to have smaller FAR upper limits, and the elasticity is estimated at around -0.035. This is likely due to the fact that large residential communities are more likely to be designed for relatively well-off households and therefore feature low-density housing.[34] The FAR upper limit also decreases slowly as one moves from city center to the suburbs, with a 50% increase in distance to city center decreasing FAR upper limit by just 1.5% to 2%. This is consistent with the common observation that the suburbs of Chinese cities are often replete with high-rise buildings.[35] In column 5 where the climatic controls are added, precision decreases a lot due to the high correlation between temperature and latitude, which renders the coefficient insignificant.

How large is the estimated effect of latitude on regulatory FAR under the sunlight policy? Our preferred specification in column 4 indicates that moving across the interquartile range of latitudes, regulatory FAR would increase by 15% all else equal. This suggests that on average, the sunlight policy is a binding constraint on urban density across China.

Table 1.4 examines whether the sunlight policy is *always* a binding constraint for density. Column 1 shows that projects on larger land parcels deviate signifcantly more from the policy, consistent with the earlier conjecture that larger residential communities are more likely to be middle-class or high-income low-density housing. In these cases, the sunlight requirement is not a binding constraint for density, as density is restricted to be lowered than that allowed by the sunlight requirement. Housing projects situated on above-median-sized parcels have an estimated β 0.2 lower than housing projects situated on below-median-sized parcels. Column 2 shows that distance to city center does not

[34] Barr and Cohen (2014) shows a similar negative correlation between plot size and FAR for commercial buildings in New York City.

[35] See, for example, https://www.wired.com/2016/12/aurelien-marechal-block-china-cities/.

significantly change the degree to which the sunlight policy shapes density across space. Projects far away from city center also appear to be constrained by the sunlight policy, consistent with the earlier hypothesis that the suburbs are also places of high-density living. Column 3 shows that FAR in larger cities deviate significantly more from what would be predicted from the sunlight policy. However, this effect disappears when controlling for GDP per capita, as column 4 indicates. Instead, richer cities appear to deviate significantly more from the policy, with housing projects in above-median GDP per capita cities having an estimated β 0.25 lower than housing projects in below-median GDP per capita cities. The main effect of GDP per capita on FAR is significantly negative, suggesting that deviation from the sunlight predictions is largely caused by projects in richer cities being subject to heavier development restrictions, rather than them violating the policy by building denser. This could be caused by richer households demanding more open space, or higher car ownership and better public transit infrastructure in richer cities lowering commuting costs and hence making distance less costly. Column 5 interacts the predicted FAR formula with an indicator for prefecture. We see that overall prefectures deviate from the sunlight policy more than counties, with β around 0.31 lower in prefectures than in counties. Column 6 adds interactions between predicted FAR and the four variables previously considered: parcel size, distance to city center, population, GDP per capita. The coefficient on the interaction between prefecture and predicted FAR drops to -0.21, suggesting that these characteristics account for some, but not all, of the differences between prefectures and counties.

One could make the argument that the observed regulatory FAR upper limits may be higher than developers' desired FAR levels (i.e., not binding). Section A.2 shows that there is a strong positive correlation between regulatory FAR and land price within cities, after a host of project-level covariates are controlled for. This is consistent with the notion that regulatory FARs tend to be binding, and these positive correlations reflect the shadow value of lifting the upper limits.

1.5 Relationship between Density and the Speed of Information Diffusion

This section starts by showing that consistent with higher residential density in the south induced by the sunlight policy, 1-km^2 grid cells with similar population densities in 1990 have diverged in population density by 2010, with cells in the south becoming significantly denser. The policy also induces differences in land prices across latitudes. The relationship between population density and the speed of information transmission is then studied. An analysis of observed dynamics of user behavior on the Local Leader Message Board reveals that people respond to issues that recently receive replies from the local government by increasing postings about the same issue. An increase in local government's reply rate to a petition topic significantly increases subsequent petitions of the same topic. The increase, however, takes time to materialize. Estimates suggest that the number of petitions reaches its peak three quarters out from the initial reply rate increase before starting to decline. This is consistent with a diffusion process of information about shifts in government attention. The response of postings to government reply is significantly faster in southern cities, where population densities are higher.

1.5.1 Latitude, Density and Land Prices

Does the sunlight policy induce higher population density and land prices in the south? I estimate the following equation:

$$y_{ic} = -\gamma \log\left(\frac{2}{3} + \frac{1}{\tan \alpha_i}\right) + \lambda_{\text{prov}(c)} + \beta_1' X_i + \beta_2' X_c + \epsilon_{ic}$$

where the unit of observation is a 1-km^2 grid cell. $-\log\left(\frac{2}{3}+\frac{1}{\tan \alpha_i}\right)$ is predicted log FAR, as previously, calculated as a function of solar elevation angle and therefore latitude. $\lambda_{\text{prov}(c)}$ is a province fixed effect. X_i is a vector of cell-level controls including log nighttime lights

in 1992 and log population density in 1990. These control for initial densities and levels of economic development before the policy was legally in place. Consequently we estimate differential changes in densities across latitudes since the early 1990s. Controlling for initial density is important as the literature has often found high persistence in population density across time, sometimes over millennia (Davis and Weinstein, 2002; Bleakley and Lin, 2012; Barjamovic et al., 2017). X_i also includes log nighttime lights in 2013 to control for potentially differential economic growth across latitudes. X_c is a vector of city-level controls including log city population in 1997, GDP per capita in 1997[36], log city population in 2012, GDP per capita in 2012 and share of land with above 15% slope within a 10-km radius from city center. Again, this controls for both initial conditions and possibly differential trends in GDP per capita and citywide population across latitudes. Alternatively, we can regard the controls in 2012 as a consequence of the sunlight policy and do not include them in the specifications. Estimates of γ turn out to be quite similar when the 2012 controls are not included.

Since the observations here are high-resolution grid cells, it is important to account for potential spatial autocorrelation in errors. I use a clustering algorithm to find contiguous "urban clusters", and cluster my standard errors at the level of these discovered spatial clusters. Section A.1.2 describes this procedure in detail. Notably, the discovered urban clusters exhibit close fits to Zipf's Law in both population and land area (see Section A.1.2). The regressions use grid cells that belong to urban clusters spanning at least 5 km^2. This excludes small patches of urban area, which may not be considered cities in the usual sense.

Table 1.5 presents the estimates. The outcome in column 1 is average residential FAR in cell, calculated as a weighted average of project FARs with project parcel size as weights. We see that the estimated coefficient is smaller than that in Table 1.3, since more weight

[36]Year 1997 falls slightly after the policy was introduced, and is the first year for which I have data for all cities.

is given to larger lots, which deviate more from the sunlight policy. The coefficient is however still sizable (above 0.5). Column 2 reports results for population density in 2010. We see that similarly dense and developed cells have diverged in population density in 2010, and cells with higher predicted FAR under the sunlight policy (i.e., lower latitudes) have become significantly denser.

A standard urban economics model with homogeneous households would predict higher land prices where population density is higher. If households spend a fixed share of their income on housing, an assumption standard since Helpman (1998), more people on one square kilometer of land implies more spending on housing within that square kilometer, which accrues to land and results in a higher price for residential land. If land is freely allocated to its most profitable use between residential and commercial uses, i.e., zoning does not distort land uses, we should expect commercial land prices to be higher in the south as well. Even if zoning causes residential and commercial land prices to be different in the same location, denser places provide greater market access to consumers, which is valuable to non-tradable goods producers in a city (e.g., restaurants, retail, etc.) since travel within cities is costly. The higher market access will be capitalized into higher prices for commercial land. Finally, to the extent that commercial floor space is elastically supplied, we should also expect to see higher commercial FARs in denser places in response to the higher commercial rent there.

These predictions are borne out in the data. Columns 3 and 4 examine residential and commercial land prices respectively. To isolate variation in land prices solely due to location, project characteristics are partialed out. In particular, I run a hedonic regression of (log) per square meter real price on cell fixed effects, log parcel size, transaction year and month-of-year dummies, land grade dummies, project type dummies and a greenfield dummy. I then use estimated cell fixed effects to represent land price in each cell. We see that both residential and commercial land are more expensive in the south, with residential land prices more responsive to latitude than commercial land prices. It is

worth noting that the share of land with above 15% slope within a 10-kilometer radius from city center is positively correlated with land price, consistent with findings in Saiz (2010) for US cities. A standard deviation increase in the share of surrounding steep slopes is associated with a 15% increase in residential land price and an 11% increase in commercial land price. In column 5, we see that denser places indeed have higher commercial FARs as well. While surrounding slopes increase residential FAR, they do not have a significant impact on commercial FAR.

1.5.2 Response of Postings to Government Replies on LLMB

To study whether people respond, either due to strategic considerations or psychological encouragements, to the type of petitions that seem to have received recent attention from the local government, I use variation in different topics being discussed on a given city's sub-board on LLMB. In an earlier paper (Jiang et al., Forthcoming), we analyzed the contents of these petitions using LDA topic models (Blei et al., 2003). Here I assign each posting to the topic that makes up the largest share of the petition. I then exclude petitions on rural issues from analysis, since we are interested in posts written by urban residents here.

Section A.5 contains highest-probability words for each topic. We see that the petitions address a wide range of issues from house expropriation, wage arrears and conflicts with property-management companies[37] to pollution, teacher compensation and pyramid schemes. Jiang et al. (Forthcoming) provides evidence that most of the postings express personal grievances rather than convey policy suggestions. I further divide the topics into grievance topics and non-grievance topics, where non-grievance topics are those that do not directly address a personal issue. These include comments on policies and compliments to government officials. Local government agencies leave publicly viewable

[37]The prevalence of this type of conflicts is again caused by the fact that the majority of Chinese urban residents live in large residential compounds that house hundreds or even thousands of families.

responses to a subset of the postings. 58.5% of postings in the sample receive replies.

It is worth noting that many of the aired grievances are not easily resolved through the legal system, either because the issue in question does not involve legal infringements (e.g., poor teacher pay and unpleasant neighborhood environment), petitioners lack the resources to resort to legal procedures (e.g., wage arrears for migrant workers) or that the justice system may not be reliable for protecting the individual (e.g., when the infringers are local government officials in housing expropriations or powerful property-management companies in property conflicts). Consequently, people seek redress for their grievances on this platform by creating publicity and trying to direct the government's attention towards their specific issue. Since the issues raised are wide-ranging, there is uncertainty around which issues the government will consider seriously. As a result, an increase in government reply rate to a topic can convey information that the government is likely to be paying attention to the issue and that the likelihood that this issue will be redressed becomes higher. We should therefore expect postings on the same issue to increase, and especially so if the topic in question is a grievance topic.

To test whether increased government reply rate to a topic increases subsequent postings on the same topic, I estimate the following equation:

$$\text{Posts}_{ict} = \gamma_{ct} + \delta_{it} + \beta \text{Reply}_{ict-1} + \epsilon_{ict}$$

where Posts_{ict} is log number of postings on topic i in city c's LLMB sub-board in quarter t. γ_{ct} are city-by-quarter fixed effects. These fixed effects allow for city-specific trends in number of petitions over time. Including them leverages variation generated by local government's differential replies to different topics within a city and a period. In other words, we are looking at whether the government's reply to a certain topic relative to other topics drives up subsequent postings on the topic relative to other topics. This guarantees that the estimated β is not confounded by city-specific trends that drive both government

replies and public engagement (e.g., popularity of the platform as an engagement tool). δ_{it} are topic by quarter fixed effects controlling for the general salience of different issues over time.

Table 1.6 presents the estimates. In column 1, city fixed effects, topic fixed effects and quarter fixed effects are included, but city-by-quarter fixed effects are not. These city-by-quarter fixed effects are included in column 2, where the estimated coefficient on last-quarter reply rate drops substantially. This suggests that the estimate in column 1 is confounded by city-specific trends capturing factors such as the platform's popularity. The estimate in column 2 is significant, and suggests that a 10% increase in reply rate to a topic will increase postings on that topic by 0.37% the next quarter. Column 3 adds topic-by-quarter fixed effects, and the coefficient remains largely unchanged.

In addition to last-quarter reply rate to the given petition topic, column 4 adds the average reply rate to all other topics in the city's sub-board last quarter. We see that this coefficient is not significantly different from 0, suggesting that the increase in postings is due to a perceived shift of government attention towards specific issues, rather than a general increase in government responsiveness. Column 5 adds an indicator for grievance topics and its interaction with last-quarter reply rate to test whether grievance topics and non-grievance topics respond differently to government reply. Grievance postings are found to be significantly more responsive to government replies. In fact, non-grievance postings appear to respond negatively to increases in reply rate, possibly due to people changing their communication tactics from policy suggestions to more outright complaints.

1.5.3 Dynamic Response to Government Replies

What does the time trajectory of response to increased government replies look like? I estimate the following distributed lag model:

$$\text{Posts}_{ict} = \gamma_c + \delta_{it} + \sum_{j=0}^{J} \beta_j \text{Reply}_{ict-j} + \epsilon_{ict} \tag{1.3}$$

Estimates for varying lag lengths from 4 to 10 are presented in Table 1.7. We see that after an increase in reply rate, postings on the same topic gradually increase and reach its peak after three quarters before starting to gradually decline.[38] This process of gradual increase likely reflects the diffusion of information about latest shifts in local government attention.

Could the observed relationship reflect a reverse causality, i.e., local governments are more likely to reply to issues that are raised more frequently? I further include leads of reply rate in Equation 1.3. The estimated coefficients adjusted to reflect responses to a 10% increase in government reply rate are plotted in Figure 1.11 for four leads and varying lengths of lags. We see that none of the leads of reply rate have significant coefficients, indicating that higher volumes of posts on a topic do not anticipate more government replies to the topic. In the three quarters after an increase in reply rate, postings gradually increase. They then start to decrease and return to baseline levels after around eight quarters.

Figure 1.13 shows the cumulative response of log number of posts to a 10% increase in government reply rate obtained from estimating the following equation:

$$\text{Posts}_{ict} = \gamma_c + \delta_{it} + \sum_{j=0}^{9} \sigma_j \Delta\text{Reply}_{ict-j} + \sigma_{10}\text{Reply}_{ict-10} + \epsilon_{ict}$$

Here an S shape of post increase becomes apparent. According to point estimates,

[38] Volumes of postings on the platform are not too high to obscure replies made a few quarters ago. An average prefecture receives around 250 postings per year during 2008-2015.

in the eight quarters after a 10% reply rate increase, postings on the same topic would increase by 5.1%.

1.5.4 Latitude and the Differential Speed of Post Increase

Does density contribute to faster information diffusion? If this is the case, we should expect posts in southern cities to respond faster to changes in government reply. To test this hypothesis, I estimate the following equation:

$$\text{Posts}_{ict} = \gamma_c + \delta_{it} + \sum_{j=0}^{J} \beta_j \text{Reply}_{ict-j} + \sum_{j=0}^{J} \eta_j \text{Reply}_{ict-j} \times \widehat{\log \text{FAR}}_c + \sum_{j=0}^{J} \text{Reply}_{ict-j} \times \sigma_j' \text{X}_c + \epsilon_{ict} \tag{1.4}$$

$\widehat{\log \text{FAR}}_c$ is "predicted log FAR" as previously defined, and is a function of latitude. In particular, $\widehat{\log \text{FAR}}_c = -\log\left(\frac{2}{3} + \frac{1}{\tan \alpha_c}\right)$, where α_c is the solar elevation angle in city c. The set of η_j's captures how latitude affects the dynamic response path of posts to government reply. To address the concern that latitude could be correlated with other characteristics than density, I also interact three sets of city-level controls with each lag of reply rate.

The first set of controls are economic controls including GDP per capita, government expenditure as a share of GDP, government social security expenditure as a share of GDP, and government low-income assistance expenditure as a share of GDP. The government expenditure measures capture the size of local government as well as weights placed on social welfare.

The second set of controls are relative shares of city residents at six education levels,[39] Controlling for education levels is important, as people with different education and skill levels might interact with each other differently and/or have different attitudes towards

[39]These are primary school, middle school, high school, 3-year college, 4-year university and master and above. These shares are tabulated from the 2005 1% census.

the government.

The third set of controls concern the city's communication infrastructure. These include the following variables normalized by city population: number of post offices, landline users, mobile phone users and broadband internet users. Faster information diffusion could be due to more extensive telecommunication networks instead of face-to-face communication facilitated by high density. I also include city's annual passengers transported and annual freight transported normalized by city population as measures of a city's general openness.

In the most saturated specification, I interact province fixed effects with each of the lags in addition to the controls, so that we only exploit differential responses to government reply within provinces.

Table 1.8 reports estimates of the η_j's. Figure 1.15 plots cumulative differential increases. The five panels in Figure 1.15 correspond to the five specifications in Table 1.8. Plotted coeffcients are appropriately scaled to reflect cumulative differential increases after a 10% increase in reply rate, between a city at the 75th percentile of latitude and a city at the 25th percentile of latitude. Positive values indicate a larger number of posts in the southern city.

Column 1 of Table 1.8 shows estimates where no controls are included. We see that in the first three quarters after an increase in government reply, cities in the south (with higher predicted log FAR) experience larger increases in posts of the same topic. In terms of magnitude, the estimates suggest that in the first two quarters after a reply rate increase, a city at the 25th percentile of latitude would experience a cumulative increase in postings that is 2.7 times as large as a city at the 75th percentile of latitude. The gap is however closing over time. After the fourth quarter, cities in the north experience larger increases in posts and therefore gradually catch up. Column 2 adds education controls. Column 3 adds GDP per capita and the government expenditure controls, and column 4 further adds the telecommunication and openness controls. We see that the

estimates only change marginally. Column 5 adds the full set of interactions between province fixed effects and reply rate lags. Estimates become larger in magnitude, but the overall pattern remains unchanged - southern cities experience larger increases in posts in the first few quarters, consistent with the hypothesis that higher density contributes to faster information diffusion.

To the extent that we worry about latitude being correlated with unobserved determinants of information transmission conditional on the controls in Table 1.8, Table 1.9 reports estimates where log predicted FAR is replaced by log observed FAR[40] as a robustness check. These estimates are potentially biased due to the endogeneity of actual FAR. Reassuringly, we see that higher density is again associated with larger responses of petitions to government attention shifts, although unlike in Table 1.8, less dense places do not appear to catch up after four quarters.

Section A.6 contains placebo tests testing for differences across longitudes, where regressions in Table 1.8 are repeated with log predicted FAR replaced by longitude. Since density does not systematically vary over longitudes, we should expect the speed of information transmission not to vary across longitudes. Table A3 suggests that this is indeed the case.

1.5.5 Direct Survey Evidence

In this subsection, I provide direct evidence on the relationship between latitude and word of mouth drawing on the China Social Governance Survey conducted by Tsinghua University researchers in 2015 (Zheng et al., 2018), which contains a host of questions on political and cultural attitudes. The survey is conducted to a nationally representative sample of respondents. I focus on the 1,362 urban respondents in my analysis, who live in 78 different cities.

[40]To be precise, observed FAR here is city-level average of residential FAR upper limit.

Table 1.10 reports results on whether the respondent heard gossip or gossiped about news concerning the economy, politics or society in the last month. Columns 1 and 3 include only latitude as an explanatory variable, while columns 2 and 4 add individual-level controls. I control for the individual's overall interest in public issues measured with the individual's frequency of consuming news about politics,[41] whether the respondent is interested in politics[42] and whether the respondent often talks about politics with family and friends.[43] The controls also include gender, education level, *hukou* status (rural or urban), age, and size of city living in.[44]

Overall, 41% of respondents indicate that they heard gossip about public issues in the previous month. A smaller share of 26% indicate that they gossiped about public issues in the previous month. Latitude is negatively associated with the probability of gossip. A five-degree increase in latitude is associated with a 3 percentage points decrease in the probability of hearing gossip and a 3.5 percentage points decrease in gossiping for similar individuals with similar levels of interest in public issues. This lends further support to the notion that higher density in the south is conducive to the spread of word of mouth.

1.6 Threats to Validity

While the results in Subsection 1.5.4 suggest that higher densities in southern cities facilitate information diffusion about local government's changes in reply priorities, there could be other reasons why postings in southern cities systematically respond faster to local government's reply rate changes, even within provinces. In this section, I discuss some of the other possible explanations, and conclude that they are not sufficiently convincing.

[41]These are fixed effects for six answers: a few times every day, once every day, a few times every week, once or twice every week, less than once every week, never.

[42]Fixed effects for four answers: completely uninterested, relatively uninterested, relatively interested, very interested.

[43]Fixed effects for three answers: often, occasionally, never.

[44]Fixed effects for population over one million, over 100,000, or under 100,000.

1.6.1 Different Attitudes towards the Government

The first alternative explanation for faster responses of postings to changes in government reply rate in southern cities within a province is that southern and northern residents may hold different attitudes towards the government, and will behave differently given the same information about shifts in government attention. To test whether this explanation is plausible, I again draw on the China Social Governance Survey.

Table 1.11 examines whether within provinces, northern and southern residents hold systematically different views about the government. Panel A concerns whether the respondent thinks the government is "solely responsible" or "primarily responsible" in guaranteeing people education, pension, jobs, healthcare and housing, as opposed to individuals securing these on their own. Specifications in odd-numbered columns include just province fixed effects, while specifications in even-numbered columns add the following individual-level controls: gender, education level, *hukou* status, age and size of city living in. We see that there is no systematic differences across latitudes in people's beliefs about government's role in any of the five areas, whether or not individual characteristics are controlled for. This is not caused by a lack of differences in opinions - among the five areas, as few as 28% of respondents think the government is responsible for providing housing, and as many as 63% of respondents think the government is responsible for providing healthcare.

Panel B shows people's trust in various entities. The outcome is whether the respondent has "complete trust" or "substantial trust" in, respectively, courts, the central government, National People's Congress, government officials, the army, police, the local government, newspapers, radio and TV programs, and social organizations. All columns include province fixed effects and individual-level controls. We see that there is no systematic difference across latitudes within provinces in terms of trust in the government institutions, although northern residents do seem slightly less likely to trust the army, and more likely to trust social organizations. Again, the lack of significance is not due to

uniform responses. The share of people who report trust ranges from 18% for government officials to 69% for the army.

1.6.2 Online Information Diffusion

Another alternative explanation for faster responses in the south to changing government priorities is that the information may be transmitted online instead of through face-to-face communication. While it seems hard to believe that people would rely solely on sending this information online without talking about it face to face, and that the regressions in Table 1.8 have included a vector of local telecommunication controls, I present evidence from the China Social Governance Survey that there are no systematic differences in internet use behavior across latitudes, so that to the extent that information travels faster in southern cities, it is likely due to offline interactions.

Table 1.12 reports the within-province relationship between individual responses to internet-related questions and the latitude of their cities. Individual characteristics are controlled for in all specifications. Column 1 concerns whether the individual can access internet through cell phone. Column 2 concerns whether the individual has other means of access to internet (broadband, optical fiber, dial-up, etc.). Column 3 is on whether the respondent uses internet for at least 30 minutes per day, and column 4 is on whether the respondent obtains political information online at least once every week. Column 5 looks at whether the respondent ever expressed political opinions online, and column 6 looks at whether the respondent uses social media. In all cases, there is no significant difference across latitudes.

1.7 Conclusion

Does urban density facilitate information and knowledge diffusion, a conjecture dating back to Marshall (1890)? Using plausibly exogenous variation induced by the sunlight

policy in China, I present evidence indicating that density does play a role in accelerating the transmission of information as evidenced by the response speed of online postings to local government's shifting attention. On a practical level, I find that a policy exogenously shifts urban density across a country with the world's largest urban population. The relationship between latitude and residential floor area ratio is clear and robust, and can be described quite well by a formula derived from developer profit maximization. This provides a useful source of variation that can be leveraged in exploring a wide range of questions regarding housing affordability, commuting costs, and urban sprawl. As the developing world undergoes rapid urbanization and city building, the way streets and buildings are laid out today could affect urban density and productivity for decades or more to come. It is therefore important to understand how urban planning practices could affect market outcomes and economic efficiency.

This paper also raises interesting questions about the governance of cities. It is not clear *a priori* whether more desirable social outcomes can be achieved if citizens are able to immediately react to policy changes. Word of mouth facilitated by urban density can either open up opportunities for designing effective policies or undermine the effectiveness of policies that are not strategy-proof. The information set of citizens mediated by density should therefore be considered in the policy-making process. On the other hand, the flow of information aided by urban density could also increase government accountability, as is suggested in Campante and Do (2014). More work is required to understand whether better knowledge of government behavior facilitated by urban density leads to increased accountability and improved governance.

If information does flow faster in denser cities, a natural question is what types of information are more sensitive to distance, and therefore benefit more from density. It is conceivable that information about rapidly changing situations - "knowledge of circumstances of the fleeting moment" in the words of Hayek (1945), or information that is tacit

and hard to codify[45] will be particularly hard to transmit over distance. Looking into these finer distinctions will enrich our understanding of the productivity benefits of cities and inform better policies. I leave this exciting topic to future studies.

[45] Juhász and Steinwender (2018) shows how the trade volume of cotton textile products with varying degrees of specification codifiability reacted differently to being connected to the global telegraph network in the 19th century.

Table 1.1: Effects of the Sunlight Policy on Building Height

	(1) High-rise	(2) High-rise	(3) High-rise	(4) High-rise	(5) High-rise	(6) High-rise
Latitude	-0.001	0.001	-0.003	-0.003		-0.004
	(0.005)	(0.004)	(0.004)	(0.004)		(0.004)
Post	0.331***	0.275***	0.271***	0.262***	0.266***	0.266
	(0.058)	(0.050)	(0.045)	(0.046)	(0.045)	(0.251)
Latitude × Post	-0.004**	-0.003*	-0.005***	-0.004***	-0.005***	-0.004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)
Slope within 10-km city radius		0.237***	0.201***	0.200***		0.197***
		(0.035)	(0.032)	(0.032)		(0.032)
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
City controls	No	Yes	Yes	Yes	No	Yes
Residence controls	No	No	Yes	Yes	Yes	Yes
Household controls	No	No	No	Yes	Yes	Yes
City FEs	No	No	No	No	Yes	No
Province trends	No	No	No	No	No	Yes
Observations	248365	240405	240091	239818	247753	239818
R-Squared	0.135	0.166	0.276	0.287	0.332	0.298
RMSE	0.371	0.364	0.339	0.337	0.326	0.334

Notes: Dependent variable is indicator that the household lives in a building with at least seven floors. Post indicates that the building was constructed in or after 1993, when the sunlight policy was written into law. City controls include 2005 levels of urban population, population density, employment, GDP per capita and share of land with above 15% slopes within a 10-km radius from city center. Residence controls include use porpose of residence (just for living or for both living and production), log number of rooms, log square footage, whether shared with other households, access to tap water, presence of kitchen, toilet and bath facilities, type of fuel and how the residence was obtained (rent or bought, SOE housing, affordable housing or commercial housing). Household controls contain log number of household members and the following characteristics of household head: gender, age, ethnicity, *hukou* registration place, type of *hukou* (rural or urban), literacy, education, whether worked for more than one hour last week, log income, source of income and marital status. Data source: 2005 census random sample. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.2: Parallel Economic Growth across Latitudes

	(1) Urban Population	(2) Employment	(3) GDP	(4) GDP pc	(5) Urban Population	(6) Employment	(7) GDP	(8) GDP pc
Year=1998 × Latitude	0.002		0.002**	0.001	0.001		-0.004	-0.005
	(0.002)		(0.001)	(0.001)	(0.003)		(0.003)	(0.003)
Year=1999 × Latitude	0.002		0.001	0.001	-0.001		-0.008	-0.006
	(0.003)		(0.002)	(0.002)	(0.005)		(0.008)	(0.007)
Year=2000 × Latitude	0.000		0.002	0.002	-0.000		-0.007	-0.005
	(0.002)		(0.002)	(0.002)	(0.007)		(0.009)	(0.009)
Year=2001 × Latitude	0.001	-0.002	0.002	0.003	0.005	-0.003	-0.008	-0.007
	(0.002)	(0.004)	(0.002)	(0.002)	(0.007)	(0.004)	(0.011)	(0.010)
Year=2002 × Latitude	-0.001	-0.003	0.004	0.006**	-0.000	0.000	-0.007	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.005)	(0.011)	(0.011)
Year=2003 × Latitude	0.004	-0.003	0.006	0.008**	-0.008	-0.003	-0.011	-0.010
	(0.005)	(0.003)	(0.004)	(0.004)	(0.009)	(0.006)	(0.013)	(0.013)
Year=2004 × Latitude	0.004	-0.003	0.008*	0.010**	-0.009	-0.004	-0.008	-0.007
	(0.005)	(0.003)	(0.004)	(0.004)	(0.009)	(0.008)	(0.015)	(0.016)
Year=2005 × Latitude	0.003	-0.006	0.009	0.012**	-0.013	-0.003	-0.005	-0.004
	(0.005)	(0.004)	(0.006)	(0.006)	(0.010)	(0.007)	(0.016)	(0.016)
City FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	No	No	No	No
Province-Year FEs	No	No	No	No	Yes	Yes	Yes	Yes
Observations	17289	12504	17964	17244	17289	12504	17964	17244
R-Squared	0.957	0.951	0.988	0.971	0.961	0.953	0.991	0.977
RMSE	0.218	0.210	0.146	0.140	0.208	0.204	0.131	0.125

Notes: City fixed effects are included in all specifications. Columns 1 to 4 include year fixed effects, while columns 5 to 8 include province by year fixed effects. Outcome variables are in logs. Each regression uses a panel that is balanced on the outcome variable. For urban population, GDP and GDP per capita, the omitted year is 1997. For employment, the omitted year is 2000. Employment data before 2000 are not available. Standard errors are clustered at province level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.3: Predictive Power of the Sunlight Policy

	DV: log FAR for Residential Project				
	(1)	(2)	(3)	(4)	(5)
Predicted log FAR	0.679***	0.628*	0.682**	0.682**	0.486
	(0.055)	(0.329)	(0.320)	(0.328)	(0.406)
Log size		-0.031***	-0.035***	-0.037***	-0.037***
		(0.003)	(0.003)	(0.003)	(0.003)
Log distance to center		-0.029***	-0.026***	-0.039***	-0.039***
		(0.004)	(0.004)	(0.004)	(0.004)
Lng-lat-cell FEs	No	Yes	Yes	Yes	Yes
Time FEs	No	No	Yes	Yes	Yes
Lot-specific controls	No	Yes	Yes	Yes	Yes
City controls	No	No	No	Yes	Yes
Climatic controls	No	No	No	No	Yes
Observations	228428	197183	197183	188402	188402
R-Squared	0.055	0.346	0.378	0.378	0.378
RMSE	0.445	0.369	0.360	0.360	0.360

Notes: These regressions concern residential development projects across China for which the land parcel is transfered between 2007 and 2016. Predicted log Floor Area Ratio is calculated as a function of latitude using implications of a simple model. See text for details. Lng-lat cells are 0.1 degrees longitude by 2 degrees latitude spatial units. Time FEs include year and month-of-year fixed effects. Lot-specific controls include whether the lot is greenfield or brown-field, type of residential building, whether the lot is transferred through allotment, contract or auction, fixed effects for land grade, log lot size, log distance to 2005 city center and 2005 nighttime lights. City controls include 2006 levels of urban population, GDP per capita, population density, employment, share of land with slopes above 15% within a 10-km radius from city center. Climatic controls include mean temperature in January, mean temperature in July, and average annual precipitation over the period of 1970-2000. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.4: Is the Sunlight Policy Always a Binding Constraint on Density?

	DV: log FAR for Residential Project					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted log FAR	0.886***	0.850**	0.933***	0.997***	0.848***	0.929***
	(0.329)	(0.331)	(0.334)	(0.330)	(0.318)	(0.325)
Parcel size, above median	-0.229***	-0.228***	-0.224***	-0.216***	-0.080***	-0.210***
	(0.035)	(0.035)	(0.034)	(0.034)	(0.006)	(0.033)
Distance to city center, above median	-0.039***	-0.004	-0.002	-0.003	-0.052***	-0.002
	(0.007)	(0.031)	(0.031)	(0.031)	(0.007)	(0.031)
Population, above median	0.037***	0.037***	-0.066	-0.020	0.015	0.025
	(0.013)	(0.013)	(0.057)	(0.059)	(0.014)	(0.064)
GDP pc, above median	-0.020	-0.020	-0.020	-0.204***	-0.026*	-0.163**
	(0.014)	(0.014)	(0.014)	(0.065)	(0.014)	(0.065)
Predicted log FAR × Parcel size, above median	-0.201***	-0.200***	-0.195***	-0.185***		-0.175***
	(0.042)	(0.042)	(0.041)	(0.041)		(0.040)
Predicted log FAR × Distance to city center, above median		0.049	0.053	0.052		0.071*
		(0.040)	(0.040)	(0.039)		(0.039)
Predicted log FAR × Population, above median			-0.141**	-0.076		0.014
			(0.072)	(0.074)		(0.080)
Predicted log FAR × GDP pc, above median				-0.254***		-0.189**
				(0.077)		(0.077)
Prefecture					-0.014	0.075
					(0.097)	(0.101)
Predicted log FAR × Prefecture					-0.309***	-0.206**
					(0.089)	(0.095)
Lng-lat-cell FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Lot-specific controls	Yes	Yes	Yes	Yes	Yes	Yes
City controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	188402	188402	188402	188402	188402	188402
R-Squared	0.368	0.368	0.368	0.368	0.368	0.369
RMSE	0.363	0.363	0.363	0.363	0.363	0.363

Notes: These regressions concern residential development projects across China for which the land parcel is transfered between 2007 and 2016. Predicted log Floor Area Ratio is calculated as a function of latitude using implications of a simple model. See text for details. Lng-lat cells are 0.1 degrees longitude by 2 degrees latitude spatial units. Time FEs include year and month-of-year fixed effects. Lot-specific controls include whether the lot is greenfield or brownfield, type of residential building, whether the lot is transferred through allotment, contract or auction, fixed effects for land grade, and 2005 nighttime lights. City controls include 2006 levels of population density and employment and share of land with slopes above 15% within a 10-km radius from city center. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.5: Latitude, Density and Prices

	(1) Resi. FAR	(2) Population	(3) Resi. Land Price	(4) Comm. Land Price	(5) Comm. FAR
Predicted log FAR	0.545***	0.772***	2.029***	1.745***	0.578**
	(0.163)	(0.202)	(0.467)	(0.470)	(0.272)
Log nighttime lights 1992	0.003	-0.069***	-0.057***	-0.004	0.021**
	(0.007)	(0.014)	(0.018)	(0.018)	(0.010)
Log nighttime lights 2013	0.072***	0.259***	0.476***	0.431***	0.174***
	(0.008)	(0.008)	(0.025)	(0.024)	(0.015)
Log population density 1990	0.093***	0.795***	0.188***	0.161***	0.101***
	(0.007)	(0.025)	(0.027)	(0.027)	(0.014)
Log city population 1997	0.033***	0.006	0.111***	0.073*	0.016
	(0.012)	(0.023)	(0.039)	(0.040)	(0.023)
Log city population 2012	-0.014	0.050***	0.062*	0.122***	-0.012
	(0.009)	(0.018)	(0.032)	(0.030)	(0.017)
GDP pc 1997	-0.021	0.152***	0.107**	0.052	-0.001
	(0.016)	(0.049)	(0.045)	(0.042)	(0.027)
GDP pc 2012	-0.046***	-0.011	0.020	0.040	-0.052**
	(0.012)	(0.025)	(0.036)	(0.032)	(0.021)
Slope within 10-km city radius	0.154***	-0.078	0.499***	0.352***	-0.043
	(0.035)	(0.049)	(0.077)	(0.089)	(0.051)
Province FEs	Yes	Yes	Yes	Yes	Yes
Observations	25439	234713	14104	11756	17810
R-Squared	0.225	0.790	0.326	0.354	0.089
RMSE	0.333	0.407	0.837	0.876	0.627

Notes: Units of observation in these regressions are 1 km^2 grid cells. Dependent variable in column 1 is log average residential FAR in the cell, calculated as a weighted average of residential project FARs with project parcel size as weights. Dependent variable in column 2 is log 2010 population density in the cell, obtained from Worldpop. Dependent variable in column 3 is residualized log residential land price in the cell, which is an estimated cell fixed effect from a hedonic regression of log per square meter real price on cell fixed effects, log parcel size, transaction year and month-of-year dummies, land grade dummies, project type dummies and a greenfield dummy. Dependent variable in column 4 is residualized log commercial land price in the cell obtained from a hedonic regression following the same procedure. Dependent variable in column 5 is log average commercial FAR in the cell, calculated as a weighted average of commercial project FARs with project parcel size as weights. Predicted log Floor Area Ratio is calculated as a function of latitude using implications of a simple model (see text for details). Standard errors are clustered at city-cluster level (see Appendix for details). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6: Reaction of Petition Posting to Government Replies

	Dependent Variable: Log Number of Posts				
	(1)	(2)	(3)	(4)	(5)
Reply rate$_{t-1}$	0.118***	0.037***	0.040***	0.042***	-0.030**
	(0.018)	(0.010)	(0.010)	(0.012)	(0.015)
Reply rate, other topics$_{t-1}$				0.022	
				(0.060)	
Grievance					-0.123***
					(0.041)
Grievance × Reply rate$_{t-1}$					0.092***
					(0.018)
City FEs	Yes	No	No	No	No
Topic FEs	Yes	Yes	No	No	Yes
Quarter FEs	Yes	No	No	No	No
Topic-Quarter FEs	No	No	Yes	Yes	No
City-Quarter FEs	No	Yes	Yes	Yes	Yes
Observations	85469	84875	84875	84875	84875
R-Squared	0.543	0.630	0.654	0.654	0.630
RMSE	0.631	0.595	0.578	0.578	0.595

Notes: Units of observation in these regressions are city-quarter-topic tuples. Topics are obtained from a 30-topic LDA model. Each post is assigned to its highest topic. Dependent variable is log number of posts about a given topic posted in a given city's Local Leader Message Board in a given quarter. Reply rate$_{t-1}$ is local government's reply rate to posts of the same topic in the previous quarter. Reply rate, other topics$_{t-1}$ is local government's average reply rate to posts of the other topics in the previous quarter. Grievance topics are described in the Appendix. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.7: Dynamic Reaction of Petition Posting to Government Replies

	Dependent Variable: Log Number of Posts			
	(1)	(2)	(3)	(4)
Reply rate	0.032	0.044	0.035	-0.003
	(0.025)	(0.029)	(0.037)	(0.047)
Reply rate$_{t-1}$	0.055***	0.053**	0.065**	0.062*
	(0.020)	(0.024)	(0.027)	(0.034)
Reply rate$_{t-2}$	0.082***	0.059**	0.063**	0.064*
	(0.020)	(0.025)	(0.030)	(0.038)
Reply rate$_{t-3}$	0.094***	0.083***	0.076***	0.096***
	(0.021)	(0.020)	(0.024)	(0.024)
Reply rate$_{t-4}$	0.086***	0.071***	0.065***	0.085***
	(0.024)	(0.023)	(0.023)	(0.027)
Reply rate$_{t-5}$		0.077***	0.074***	0.065***
		(0.026)	(0.025)	(0.025)
Reply rate$_{t-6}$		0.073***	0.069***	0.072***
		(0.025)	(0.022)	(0.026)
Reply rate$_{t-7}$			0.051	0.056*
			(0.032)	(0.032)
Reply rate$_{t-8}$			0.026	0.009
			(0.030)	(0.028)
Reply rate$_{t-9}$				0.027
				(0.033)
Reply rate$_{t-10}$				-0.009
				(0.032)
City FEs	Yes	Yes	Yes	Yes
Topic-Quarter FEs	Yes	Yes	Yes	Yes
Observations	26782	19579	14858	11505
R-Squared	0.598	0.622	0.637	0.651
RMSE	0.565	0.564	0.563	0.560

Notes: Units of observation in these regressions are city-quarter-topic tuples. Topics are obtained from a 30-topic LDA model. Each post is assigned to its highest topic. Dependent variable is log per capita number of posts about a given topic posted in a given city's Local Leader Message Board in a given quarter. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.8: Differential Speed of Petition Reaction to Government Replies

	Dependent Variable: Log Number of Posts				
	(1)	(2)	(3)	(4)	(5)
Reply rate × Predicted log FAR	0.201	0.208	0.117	0.218	2.129***
	(0.153)	(0.167)	(0.184)	(0.222)	(0.719)
Reply rate$_{t-1}$ × Predicted log FAR	0.355***	0.376**	0.303*	0.515***	0.879
	(0.131)	(0.150)	(0.159)	(0.186)	(0.678)
Reply rate$_{t-2}$ × Predicted log FAR	0.148	0.057	0.086	0.153	0.844
	(0.135)	(0.154)	(0.165)	(0.193)	(0.717)
Reply rate$_{t-3}$ × Predicted log FAR	0.057	0.123	0.166	0.272	0.710
	(0.131)	(0.154)	(0.163)	(0.189)	(0.670)
Reply rate$_{t-4}$ × Predicted log FAR	-0.305**	-0.416***	-0.277*	-0.361*	0.713
	(0.130)	(0.148)	(0.161)	(0.185)	(0.671)
Reply rate$_{t-5}$ × Predicted log FAR	-0.231*	-0.287*	-0.286*	-0.232	1.121*
	(0.130)	(0.148)	(0.157)	(0.178)	(0.634)
Reply rate$_{t-6}$ × Predicted log FAR	-0.304**	-0.434***	-0.359**	-0.203	-0.995
	(0.134)	(0.154)	(0.165)	(0.185)	(0.663)
City FEs	Yes	Yes	Yes	Yes	Yes
Topic-Quarter FEs	Yes	Yes	Yes	Yes	Yes
Education interactions	No	Yes	Yes	Yes	Yes
Economic interactions	No	No	Yes	Yes	Yes
Communication interactions	No	No	No	Yes	Yes
Province FE interactions	No	No	No	No	Yes
Observations	19317	19252	18538	18176	18176
R-Squared	0.623	0.626	0.613	0.618	0.627
RMSE	0.565	0.564	0.566	0.566	0.562

Notes: Units of observation in these regressions are city-quarter-topic tuples. Topics are obtained from a 30-topic LDA model. Each post is assigned to its highest topic. Dependent variable is log number of posts about a given topic posted in a given city's Local Leader Message Board in a given quarter. Predicted log Floor Area Ratio is calculated as a function of latitude using implications of a simple model (see text for details). Coefficients of reply rate and its lags are omitted. Education interactions are reply rate and its lags interacted with shares of six education levels of city residents. Economic interactions are reply rate and its lags interacted with city's GDP per capita, fiscal expenditure to GDP ratio, social security spending to GDP ratio and social assistance to GDP ratio. Communication interactions are reply rate and its lags interacted with the following city-level variables normalized by population: annual passengers transported, annual freight transported, number of post offices, landline users, mobile phone users, broadband internet users. Province FE interactions are reply rate and its lags interacted with province fixed effects. Standard errors are clustered at city-topic level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.9: Differential Speed of Petition Reaction to Government Replies Using Observed FAR

	Dependent Variable: Log Number of Posts				
	(1)	(2)	(3)	(4)	(5)
Reply rate × log FAR	0.126**	0.145***	0.171***	0.204***	0.224**
	(0.054)	(0.054)	(0.059)	(0.066)	(0.102)
Reply rate$_{t-1}$ × log FAR	0.013	0.022	0.052	0.077	0.090
	(0.047)	(0.048)	(0.054)	(0.058)	(0.089)
Reply rate$_{t-2}$ × log FAR	0.063	0.075	0.052	0.088	0.206**
	(0.046)	(0.048)	(0.055)	(0.061)	(0.087)
Reply rate$_{t-3}$ × log FAR	0.071	0.127***	0.103**	0.171***	0.190**
	(0.044)	(0.046)	(0.052)	(0.058)	(0.088)
Reply rate$_{t-4}$ × log FAR	0.015	0.064	0.023	0.043	0.064
	(0.048)	(0.050)	(0.054)	(0.058)	(0.085)
Reply rate$_{t-5}$ × log FAR	-0.043	-0.019	-0.042	-0.001	-0.046
	(0.049)	(0.049)	(0.053)	(0.059)	(0.081)
Reply rate$_{t-6}$ × log FAR	0.060	0.087*	0.040	0.121**	0.091
	(0.047)	(0.047)	(0.050)	(0.056)	(0.080)
City FEs	Yes	Yes	Yes	Yes	Yes
Topic-Quarter FEs	Yes	Yes	Yes	Yes	Yes
Education interactions	No	Yes	Yes	Yes	Yes
Economic interactions	No	No	Yes	Yes	Yes
Communication interactions	No	No	No	Yes	Yes
Province FE interactions	No	No	No	No	Yes
Observations	19211	19146	18432	18070	18070
R-Squared	0.622	0.625	0.613	0.618	0.626
RMSE	0.565	0.564	0.566	0.566	0.562

Notes: Units of observation in these regressions are city-quarter-topic tuples. Topics are obtained from a 30-topic LDA model. Each post is assigned to its highest topic. Dependent variable is log number of posts about a given topic posted in a given city's Local Leader Message Board in a given quarter. Log Floor Area Ratio is log of city-level average residential FAR. Coefficients of reply rate and its lags are omitted. Education interactions are reply rate and its lags interacted with shares of six education levels of city residents. Economic interactions are reply rate and its lags interacted with city's GDP per capita, fiscal expenditure to GDP ratio, social security spending to GDP ratio and social assistance to GDP ratio. Communication interactions are reply rate and its lags interacted with the following city-level variables normalized by population: annual passengers transported, annual freight transported, number of post offices, landline users, mobile phone users, broadband internet users. Province FE interactions are reply rate and its lags interacted with province fixed effects. Standard errors are clustered at city-topic level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.10: Survey Evidence on Word of Mouth

	Experience in the last month			
	(1) Heard Gossip	(2) Heard Gossip	(3) Gossiped	(4) Gossiped
Latitude	-0.005	-0.006**	-0.006**	-0.007***
	(0.003)	(0.002)	(0.003)	(0.002)
Individual controls	No	Yes	No	Yes
Mean of DV	0.410	0.411	0.263	0.266
Observations	1362	1324	1362	1324
R-Squared	0.003	0.187	0.005	0.191
RMSE	0.491	0.448	0.439	0.401

Notes: Dependent variable is whether the respondent heard gossip or gossiped about news concerning the economy, politics or society in the last month. Individual controls include respondent's frequency of consuming news about politics, whether the respondent is interested in politics, whether the respondent often talks about politics with family and friends, gender, education level, *hukou* status, age, and size of city living in. Data are for urban respondents in the 2015 China Social Governance Survey. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.11: Attitudes towards Government not Systematically Different Across Latitudes

	Government should be responsible for									
	(1) Education	(2) Education	(3) Pension	(4) Pension	(5) Jobs	(6) Jobs	(7) Healthcare	(8) Healthcare	(9) Housing	(10) Housing
Latitude	0.009	0.009	-0.001	0.002	-0.002	0.005	-0.021	-0.022	-0.017	-0.013
	(0.016)	(0.017)	(0.018)	(0.017)	(0.020)	(0.019)	(0.017)	(0.017)	(0.016)	(0.016)
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Mean of DV	0.618	0.623	0.541	0.545	0.373	0.374	0.629	0.631	0.280	0.278
Observations	1362	1324	1362	1324	1362	1324	1362	1324	1362	1324
R-Squared	0.047	0.056	0.045	0.056	0.037	0.057	0.036	0.061	0.044	0.063
RMSE	0.479	0.477	0.492	0.491	0.479	0.476	0.479	0.474	0.444	0.440
	Have Trust in									
	(1) Courts	(2) Central Govt	(3) NPC	(4) Officials	(5) Army	(6) Police	(7) Local Govt	(8) Newspapers	(9) Radio and TV	(10) Social Org.
Latitude	0.014	-0.002	-0.000	0.002	-0.020*	-0.001	-0.000	0.001	0.016	0.019**
	(0.018)	(0.015)	(0.016)	(0.013)	(0.012)	(0.016)	(0.013)	(0.018)	(0.013)	(0.009)
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of DV	0.440	0.674	0.602	0.184	0.687	0.440	0.282	0.222	0.278	0.103
Observations	1324	1324	1324	1324	1324	1324	1324	1324	1324	1324
R-Squared	0.048	0.032	0.051	0.049	0.054	0.041	0.033	0.030	0.044	0.051
RMSE	0.491	0.467	0.484	0.383	0.457	0.493	0.448	0.415	0.444	0.300

Notes: Dependent variable in Panel A is whether the respondent believes that the government should be "solely responsible" or "primarily responsible" in the listed areas. Dependent variable in Panel B is whether the respondent has "complete trust" or "substantial trust" in the listed entities. Individual controls include gender, education level, *hukou* status, age, and size of city living in. Data are for urban respondents in the 2015 China Social Governance Survey. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.12: Internet Use not Systematically Different Across Latitudes

	(1) Mobile Access	(2) Other Access	(3) Frequent Use	(4) Get Info	(5) Air Opinions	(6) Social Media
Latitude	-0.004	-0.010	0.001	0.011	0.003	0.016
	(0.008)	(0.017)	(0.013)	(0.014)	(0.009)	(0.009)
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of DV	0.625	0.644	0.483	0.393	0.122	0.572
Observations	1324	1324	1324	1324	1324	1324
R-Squared	0.511	0.298	0.429	0.330	0.095	0.530
RMSE	0.343	0.407	0.383	0.405	0.315	0.344

Notes: The dependent variable is whether the respondent has internet access from cell phone in column 1 , through other means (broadband, optical fiber, dial-up, etc.) in column 2, whether the respondent uses internet for at least 30 minutes per day in column 3, whether the respondent obtains political information online at least once every week in column 4, whether the respondent ever expressed political opinions online in column 5, and whether the respondent uses social media in column 6. Individual controls include gender, education level, *hukou* status, age, and size of city living in. Data are for urban respondents in the 2015 China Social Governance Survey. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

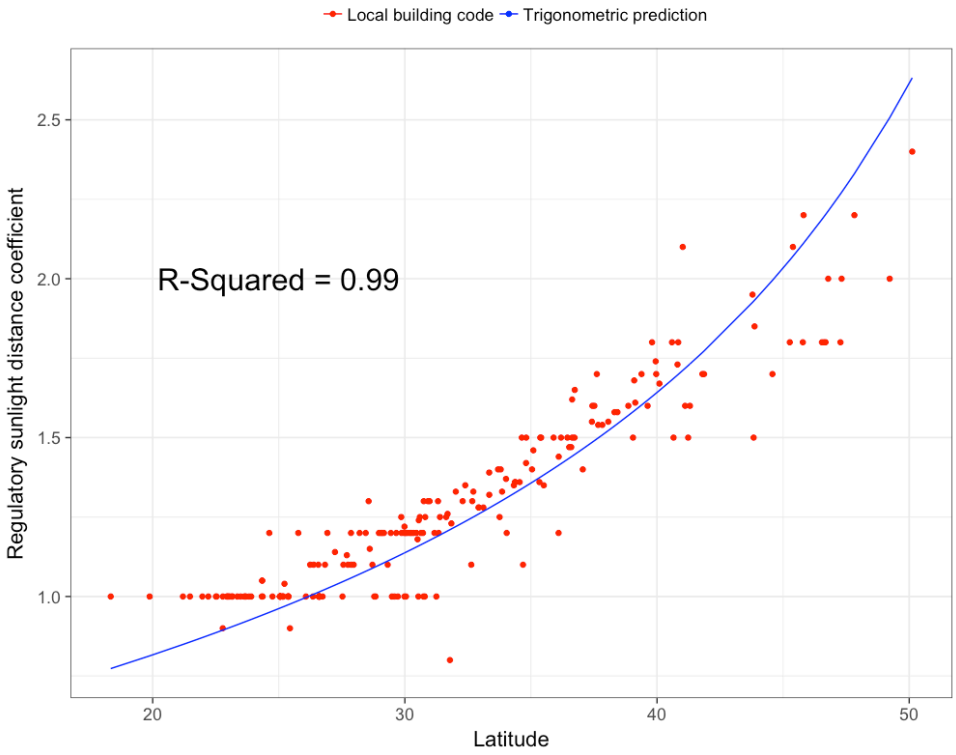


Figure 1.1: Sunlight Distance Coefficients and Latitude

Notes: Y-axis is the sunlight distance coefficient specified in a prefectural building ordinance for residential buildings below seven stories in new urban areas. X-axis is latitude of the prefectures. The blue curve shows the theoretical minimum distance required between buildings to satisfy the sunlight policy at each latitude, equal to cotangent of the solar elevation angle at each latitude at 11 am on January 20, the day specified by the policy.

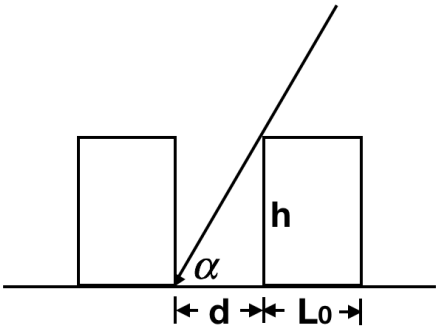


Figure 1.3: Solar Elevation Angle and Building Height

Notes: This figure shows the trade-off between building height and distance between buildings under a given solar elevation angle α.

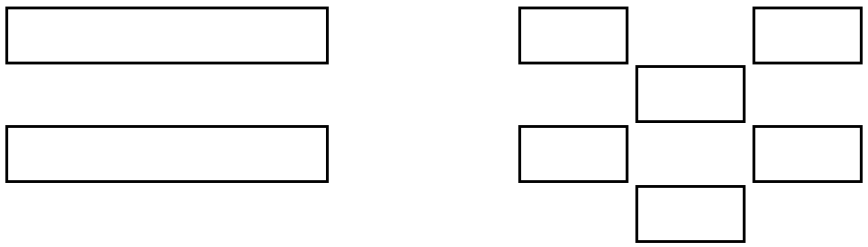


Figure 1.5: Illustrations of Building Layout

Notes: This figure shows that the two layouts of buildings achieve the same amount of floor space.

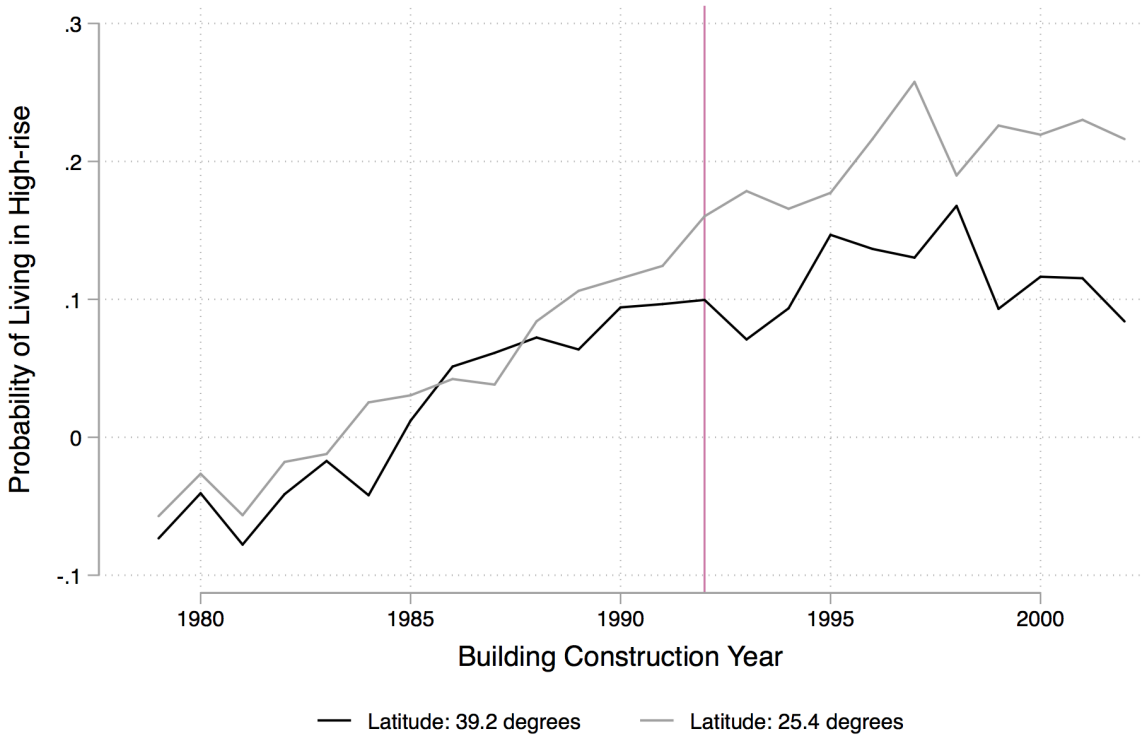
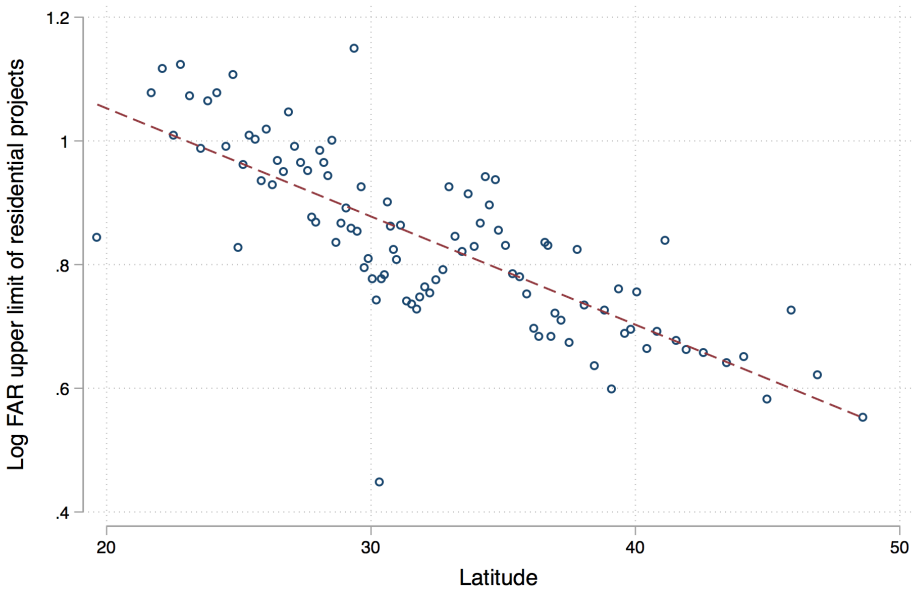
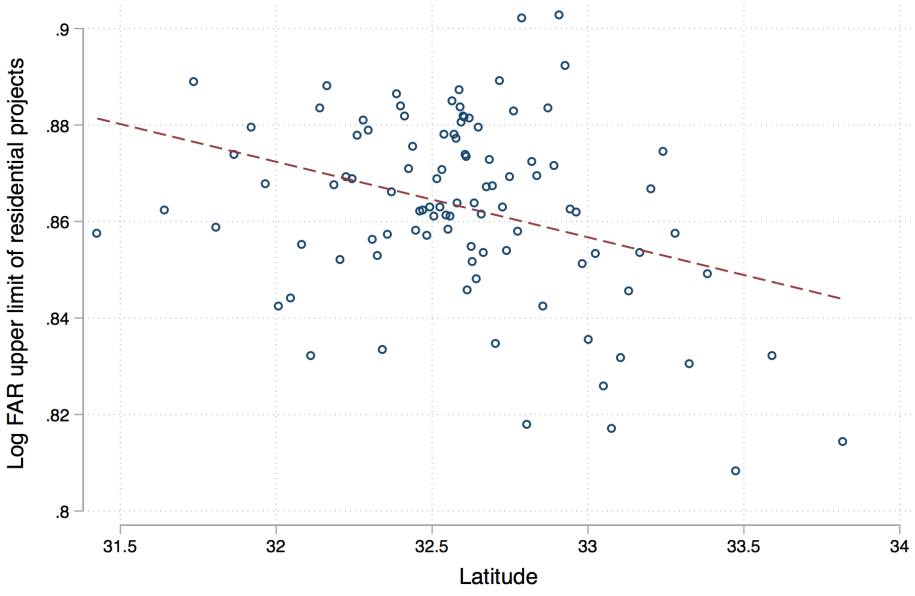


Figure 1.7: High-rises in the North and the South, by Construction Year

Notes: This figure shows estimated probability of a household living in a high-rise conditional on construction year of the building in a northern city (75th percentile of latitude) and a southern city (25th percentile of latitude). These estimates are obtained by combining year fixed effects and the interactions of year fixed effects and latitude in a household-level regression using 2005 census data where city fixed effects, residence controls and household controls are included. See text for details of the regression and the controls. Red vertical line shows the year preceding the sunlight policy.



(a) No controls



(b) Within longitude-latitude cells and full controls

Figure 1.9: Binned Scatterplot of Regulatory Floor Area Ratio and Latitude

Notes: These figures are binned scatterplots of the latitude and log regulatory FAR of individual residential lots from across China. No controls are included in Panel A. Panel B shows latitude and regulatory FAR residualized within 2 degrees latitude by 0.1 degrees longitude cells and over year fixed effects, month-of-year fixed effects, nighttime light in 2005, a host of lot-level controls and a host of city-level economic controls. See details in text.

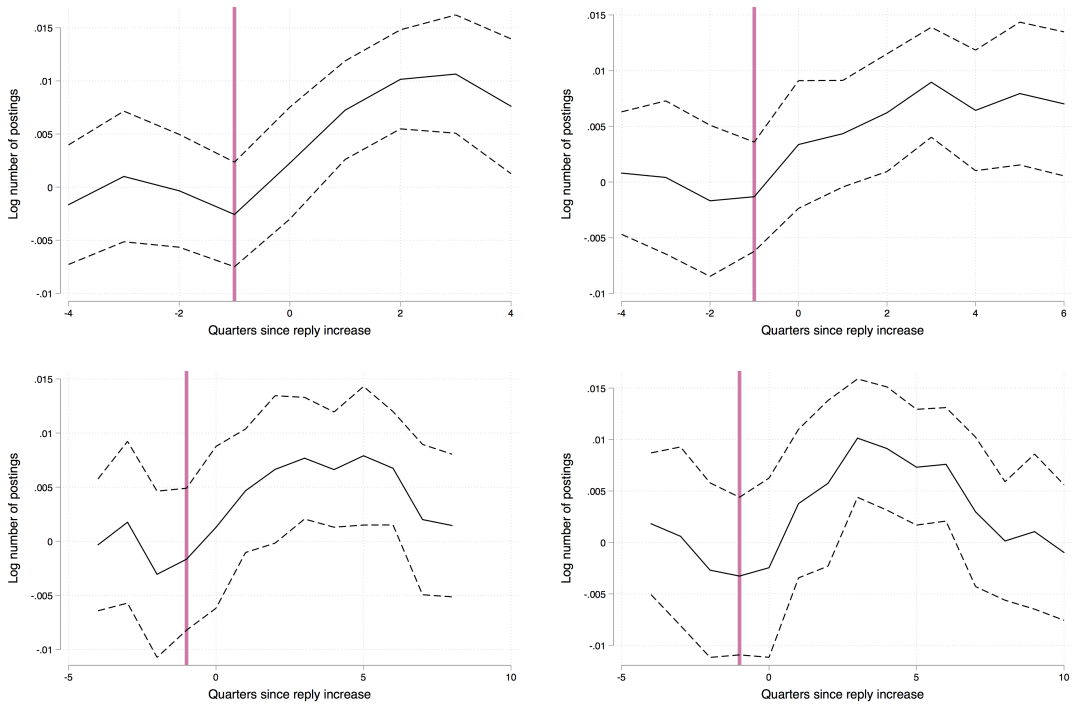


Figure 1.11: Dynamic Response of Postings to a 10% Increase in Government Reply Rate

Notes: These figures show coefficients and their 95% confidence intervals estimated from distributed lag models of the dynamic response of number of posts on a topic to a 10% increase in local government reply rate to the same topic. City and topic-by-quarter fixed effects are included in all regressions. The four panels show estimates using 4, 6, 8 and 10 lags of reply rate respectively. Four leads of reply rate are also included in each model. Red vertical lines mark quarter preceding a reply rate increase.

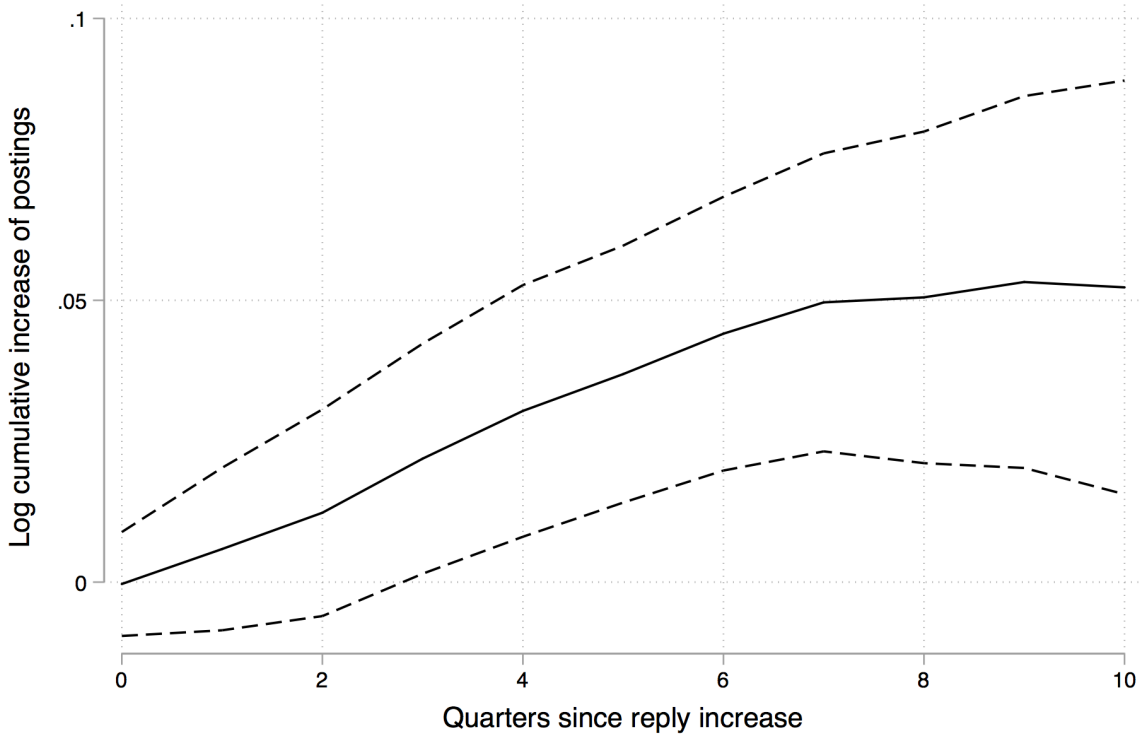


Figure 1.13: Cumulative Increase in Postings after a 10% Increase in Government Reply Rate

Notes: This figure shows the cumulative increase in log number of postings on a topic after a 10% increase in local government reply rate to the same topic. City fixed effects, topic-by-quarter fixed effects and four leads of reply rate are included in the regression. Dashed lines show 95% confidence intervals.

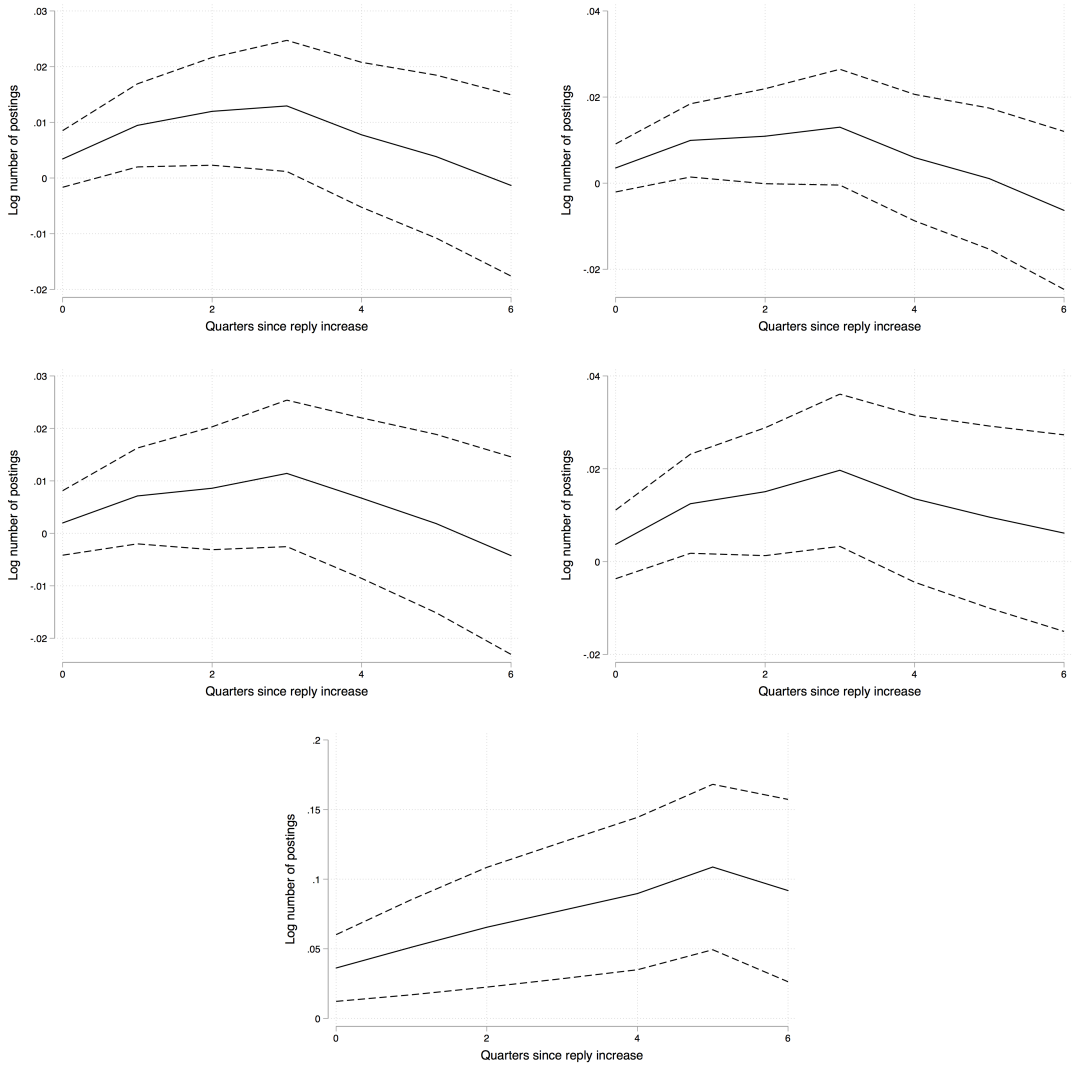


Figure 1.15: Cumulative Differential Responses of Postings to a 10% Increase in Reply Rate

Notes: These figures show the cumulative difference in log number of postings between the 75th percentile and the 25th percentile of latitude each quarter after a 10% increase in government reply rate. Positive values indicate larger numbers of posts in southern cities. The five panels correspond to estimates of the five specifications in Table 1.8. City and topic-by-quarter fixed effects are included in all regressions. Dashed lines show 95% confidence intervals.

Chapter 2

From Internet to Social Safety Net: The Policy Consequences of Online Participation in China

2.1 Introduction

Effective use of digital technologies is an essential part of modern governance. In the past two decades, national and local governments around the world have set up new Internet-based interfaces to interact with their citizens. In countries as diverse as Japan, United States, China, Colombia, and Ukraine, online participation (also known as e-participation) has become a common method for citizens to contact political authority and articulate their issues and grievances (United Nations, 2016, p.56). These developments have also triggered an intense debate among scholars about the nature and policy ramifications of online participation. While optimists view online participation as having the potential to expand civic engagement, enhance responsiveness of public institutions, and promote transparency and equity in policy making (Bimber, 1998; Castells, 1996), others remain skeptical about its ability to benefit underrepresented constituencies and its effectiveness in inducing broad policy changes(Chadwick and May, 2003; Lindner and Riehm, 2011).

In this study, we contribute new evidence to this debate by analyzing online participation in China. After almost two decades of rapid expansion of the Internet, China now hosts the largest Internet-using population in the world.[1] The Chinese government has

[1]According to China Internet Network Information Center (CNNIC), the number of Internet users in China is estimated to be over 710 million as of July 2016.

been very active in creating various types of e-participation initiatives, including electronic petition platforms, online public consultation, and government-sponsored social media accounts.[2] However, the question remains whether these initiatives are just for window-dressing purposes, or whether they have genuine influence over government policies. To answer this question, our analysis focuses specifically on the *Local Leader Message Board* (LLMB), a major online petition forum that allows citizens to directly register complaints to party and government leaders in their localities on the Internet. Launched by China's central media since 2008, the LLMB exemplifies a new class of government-sponsored electronic participation platforms that are now being adopted worldwide (see Table B1 for a list of similar institutions in other countries/regions). Compared to the more conventional participatory institutions, this platform offers an unusually cheap and transparent way for citizens to communicate their issues and concerns to the authority. As a result, we hypothesize that it would be especially attractive to lower-class citizens, who tend to be discriminated against in the "normal" political process and are usually most sensitive to changes in participation costs. Based on a large body of research suggesting that lower-income groups tend to hold stronger preference for redistribution (Alesina and Ferrara, 2005; Meltzer and Richard, 1981), we further postulate that expanded online participation will shift government policies towards placing greater emphasis on social welfare and redistributive issues, as the voices of the poor become better heard.

To evaluate these hypotheses, we first examine the content of petitions filed on the LLMB platform. We manually code issues and user backgrounds for a random sample of petitions, and then use a non-parametric content analysis method developed by Hopkins and King (2010) to estimate their distributions within the entire body of over 900,000 petitions. Our analysis shows that a substantial share of the LLMB petitions concern personal problems that are most likely to be experienced by lower-class citizens, including

[2]As of 2016, China is ranked at 22 among the top 50 performers in e-participation(United Nations, 2016), a standing on par with Denmark and higher than Luxembourg (43), Portugal (50), and Iceland (50).

employment, neighborhood environment, and land-taking compensation. Consistent with this, we also find that a sizable share of the petitions are originated from rural and suburban areas, which are inhabited primarily by individuals who belong to the lower strata of the Chinese society.[3]

We then investigate how expanded online participation affected substantive policy priorities of local governments. To construct a consistent measure of government policies that can be compared across time and space, we apply natural language processing techniques on over 4,400 Government Work Reports (GWRs), which are comprehensive policy blueprints published annually by the Chinese government at various levels, and estimate the relative proportions of different "topics" (i.e., clusters of words) as a measure for government policy priorities. Our baseline results show that cities that receive a larger number of online petitions in a year tend to devote significantly higher proportions of government reports in the following year to a topic on social welfare, which broadly includes issues such as low-income subsidies, unemployment protection, and medical assistance. We also find that the policy effect appears to be mainly driven by petitions filed by *rural* residents concerning issues related to their *pocketbook conditions*. We subject these results to a number of robustness checks and adopt several different strategies to address the problem of endogeneity between online petition and government policy. Furthermore, leveraging data from the Minimum Living Standard Guarantee Scheme (*dibao*), a key low-income assistance program targeting the poor, we show that those changes in stated policy priorities are not merely empty talk, but are accompanied by substantive improvements in the coverage of the *dibao* program.

This paper contributes to the literature on e-governance. Researchers have investigated the nature and characteristics of different e-governance techniques (e.g., Chadwick

[3]In this article, we conceptualize the lower class in the Chinese context as including farmers and low-skill/migrant workers who usually live in rural areas or city outskirts (i.e., towns/suburban areas). This is consistent with the prevailing views about the social hierarchy in contemporary China (e.g., Goodman, 2014).

and May, 2003; Norris and Reddick, 2012; Torres et al., 2006), the conditions under which certain e-governance techniques are adopted (e.g., Royo et al., 2013; Wong and Welch, 2004), and the relationship between citizens' attitudes and their participatory experiences (e.g., Kim and Lee, 2012). However, there is still little systematic evidence on the critical question of whether online participation can have any substantive impact in government policies. Some research on advanced democracies suggests that e-government initiatives typically emphasize efficient service delivery over expanding participation opportunities for citizens (Chadwick and May, 2003; Norris and Reddick, 2012). Others analyze cross-national data and find that development in e-participation are not strongly associated with reduction in corruption perception or improvement in quality of government (e.g., Linde and Karlsson, 2013). Leveraging more fine-grained subnational data on online participation and government policies, our study offers evidence that e-participation can indeed pressure local governments to devote more attention to social welfare issues even in the absence of electoral accountability. These findings lend support to the view that the Internet and information technologies can be a tool of empowerment for citizens in the digital age (Dutton, 2009).

2.2 Background

2.2.1 Online Participation in China and the *Local Leader Message Board*

The rapid expansion of the Internet in China since the 1990s and especially during the first decade of the 21st century has brought profound changes to the way citizens and governments interact. The emergence of online forums, weblogs, and many other types of virtual communities have facilitated lively discussion of public affairs and provided citizens with new channels to articulate their problems and interests (Yang, 2009). At the same time, the government has also undertaken systematic efforts to expand its online

presence. The word "E-government" first appeared in a major party policy document at the 16th Party Congress in 2002, and a series of regulations were subsequently promulgated to promote the development of central and local government websites, with a special emphasis on information availability and ease of access. As of 2008, virtually all local governments at or above the county-level had set up their own websites, and many used these websites regularly for disseminating information and delivering various types of services (Pan, 2017; Stromseth et al., 2017).

To understand the patterns and policy consequences of online participation in China, this article focuses on a major government-sponsored petition platform named the *Local Leader Message Board* (LLMB). The LLMB was created in 2008 by the official website of China's central media with the goal of providing an integrated portal for citizens to contact leading government officials in their localities (for a snap shot of the interface, see Figure B1 in Online Appendix). Ever since its launch, the platform has become a popular venue for citizens to report issues and register complaints to local authorities. As of August 2016, the LLMB has received over 900,000 petitions from all mainland provinces, making it the most heavily used platform in China today.[4]

LLMB hosts two message boards for each subnational unit at or above the county level (i.e., 33 provinces, 333 cities, and over 2,000 counties), one for the head of the party and the other for the head of the government. Participants can post their petitions to these message boards either anonymously or with a registered account. Registration requires only a functional cell phone number or an email address, and will allow the users to customize the names that appear on their petitions. After a petition is filed and before

[4]The LLMB is not the only online petition platform available in China. Inspired by the LLMB, many local governments later also developed similar platforms to receive and process petitions (for related studies, see Chen and Kung (2016); Distelhorst and Hou (2017)). However, because of the concern for censorship and data availability issues (many local platforms do not make petition data publicly available), we only focus on the LLMB for this study. In Online Appendix B.4 we compare LLMB petitions with petitions at a select set of local platforms on which we can get complete data. We find that the volumes of petitions from the two sources are highly correlated. This suggests that petitions at LLMB can be seen as broadly representative of the general patterns of online petitions in China.

it is publicly posted, the website will conduct a quick check on the content. While some censorship does happen at this stage, our interviews with LLMB staff suggest that the criteria are much more lenient than those used for social media sites.[5] According to one member of the management team, petitions are typically only blocked if they contain vulgar languages or explicit comments/criticisms of top political leaders, and "99.9% of the petitions are displayed exactly the way they were written" (Personal interview, Beijing, December 2016).

Insofar as responsiveness is concerned, several features of the LLMB deserve special mentioning. First, in contrast to locally operated websites, in which local authorities have a direct control over the content to be displayed, the LLMB is operated by a central agency that typically has little direct interest in helping local authorities cover up their problems. This means that local governments will not be able to simply bury any undesirable issues through censorship. In addition, the website also employs several measures to increase the publicity of petitions and put pressure on local government to respond. For example, all petitions and government replies are publicly visible to all users once posted; this not only makes it easier for local residents with common grievances to find each other and organize, but can also help local leaders' political superiors to learn about citizens' opinions about their subordinates. Occasionally, the LLMB's own news team will look for newsworthy materials from the petitions and conduct follow-up investigations that will be published on the central media's website. All these features—the relative independence of the operator, the transparent design of the website, and the high publicity of the platform—may give local officials an incentive to be attentive to citizen demands made on the platform.

[5] This is in part because, unlike an online forum, LLMB users only interact with the government but not with each other; the lack of horizontal communication makes it less pressing to censor posts for the purpose of preventing collective actions.

2.3 Hypotheses

In this section, we develop hypotheses regarding (1) the type of citizens who would be drawn to the LLMB and (2) the potential policy consequences associated with online participation. First, we hypothesize that *lower-class citizens are more likely to use these online platforms than those from a better-off background*. The main function of the LLMB is facilitate citizens' *articulation of personal grievances* to local political leaders. According to the classical resource-mobilization theory of political participation, individualized interest articulation is a rather resource-demanding form of participation because there is no standardized menu for actions and the outcome depends largely on a citizen's own ability and initiatives (Verba and Nie, 1972; Marien et al., 2010). This observation also holds in the context of China. Those who can afford living in major cities (i.e., closer to where the higher-level governments reside) or offering bribes to build personal connections with government officials are usually in a better position to influence government than those who lack such resources.[6] By drastically lowering the time and costs needed for long-distance communication, the LLMB can help improve rural residents' ability to contact higher-level governments located in distant urban centers. Its integrated and simplified interface also helps lower the knowledge barriers for navigating complex bureaucratic systems. Furthermore, the anonymous and impersonal nature of the website means that it would be much harder for officials to make differential treatment based on participants' political connections or other personal characteristics (Ward et al., 2003). In contrast to the upper-class citizens, whose background may give them an advantage in contacting government officials, anonymity is especially attractive to lower-class citizens, who are most likely to be discriminated against for their background.

[6]Consistent with this view, existing research shows that the traditional off-line participation channels in China are disproportionately used by those of higher political and socioeconomic status. Guo (2007), for example, finds that members of the Chinese Communist Party (CCP) and various "mass organizations" have a higher propensity to utilize the official channels in political participation than do non-members. Tsai and Xu (Forthcoming) show that citizens with political connections are more likely to contact the authorities with complaints about public services.

It is also worth noting that, unlike many other developing countries, lower-class citizens in China do have the access to the necessary Internet *infrastructure* for effective online participation. Since the early 2000s, the government and state-owned telecommunication companies have made systematic efforts in building up Internet-related infrastructure (e.g., electricity, telephone lines, etc) in remote, rural areas (Oreglia, 2015). As of 2009, Internet access was available in over 90% of the administrative villages.[7] The recent advancement in mobile telecommunication technologies (i.e., 3G and 4G networks) has led to a further expansion of Internet users in rural areas. One estimate suggests that over 87% of the rural Internet users in China today access Internet primarily through smart phones (CNNIC, 2016). The wide availability of relatively inexpensive Internet access makes rural and lower-class citizens an influential group in China's cyberspace. Although direct evidence from political sphere remains limited, indications can be found from patterns of Internet usage in other areas: Research has shown, for example, that rural residents in China are highly active in using the Internet for *commercial* purposes (Liu et al., 2015). According to a study by McKinsey, online retail channels created more new spending in China's underdeveloped regions than in developed ones (Dobbs et al., 2013). If commercial activities are of any guidance, they suggest that the rise of online participation opportunities such as the LLMB will also likely attract more new participants from the less prosperous areas.

If the LLMB can indeed help reduce barriers for lower-class participation, a question that naturally follows is: how would that affect government policies? A large literature on political economy suggests lower-class citizens often hold stronger preferences for redistribution (e.g., Acemoglu and Robinson, 2005; Meltzer and Richard, 1981; Hibbs et al., 1982). Thus, our second hypothesis is that that *government policies will become more redistributive as the voices of the disadvantaged become better heard*. Although govern-

[7]Ministry of Industry of Information Technology, https://goo.gl/CBTGrs. Although ownership of personal computer in rural areas is still not as common as in urban areas, the presence of Internet cafes and other public access venues provides inexpensive alternatives.

ment officials in China are often not held electorally accountable to their citizens, existing studies suggest several reasons why citizen participation can still lead to policy changes even in the absence of electoral institutions. Participation may, for example, reveal important information about citizen preferences and local conditions (Distelhorst and Hou, 2017; Lorentzen, 2013), which can be used by an official's superior to evaluate his/her performance. Compared to electoral democracies, the Chinese regime may be even more dependent on information conveyed through channels such as the Internet because of the underdevelopment of more conventional democratic institutions. Moreover, responsiveness may also be driven by officials' fear of citizens' collective actions (Chen and Kung, 2016). To the extent that online grievances are important signals of mass dissatisfaction, authorities may be motivated to make substantive policy concessions in response to these grievances as a way to preempt more disruptive offline actions.

Several pieces of anecdotal evidence suggest that local governments do take online petitions from the LLMB seriously. Shortly after the platform was launched, many local governments formed partnership with the platform and set up specialized agencies to process petitions and regularly report key information from online petitions to high-level decision-makers.[8] When drafting policy documents, such as the Government Work Reports (see below), the central and local authorities have also shown a willingness to seek and incorporate suggestions from online platforms.[9] Moreover, many local politicians use this platform to publicly engage with their constituencies: As of 2015, over 50 provincial party secretaries and governors, the highest-ranking regional leaders in China, have written personal replies to citizen petitions at the LLMB, along with many more city- and county-level leaders.[10]

[8]Personal interviews, 2017. See Online Appendix B.17 for more information about the petition-policy linkage.

[9]The central government, for example, initiated a campaign entitled "I spoke to the Premier" to encourage netizen input to the drafting of the Central Government Work Report in 2014. Many local governments have also made their own initiatives. See *Xinhua News*, https://goo.gl/d1oDQP, and *Shenzhen News*, https://goo.gl/qwS8SG.

[10]On average, about 60% of the petitions on LLMB have received replies from local government

2.4 Analyzing LLMB Petitions: Issues and User Backgrounds

We scraped all the publicly available petitions from the website (~900,000), along with a rich set of information for each petition, such as (user-classified) subject matter, time of posting, the identity of the leader to whom the petition was directed, whether a government reply has been made, and so on. We began with a descriptive analysis of the range of issues raised on the platform as well as petitioners' areas of origins. To do so, we took a random sample of 3,500 petitions and hired two research assistants (RAs) to read through and assign to each petition one of 14 issue labels and 3 location labels based on the content of the petition.[11] Since not all petitions contain sufficient information for us to make unambiguous coding decisions, sample statistics based on coded petitions alone are likely to be biased.[12] To remedy this, we adopted a non-parametric method developed by Hopkins and King (2010) to infer the overall distribution of issues and locations in the entire body of petitions based on the coded sample. The basic intuition of this method is that the frequencies of different words in a corpus can be expressed as the product of (1) the word frequencies in different issue categories and (2) the relative shares of these categories. Using the hand-coded information of word frequencies in different categories (i.e., labels) from the training sample as a proxy for quantity (1), we can back out the label composition that is most likely to generate the aggregate word frequencies observed in the entire body of petitions (i.e., quantity (2)).[13] We provide more details on the

agencies.

[11] In creating the issue groups, we consulted the output from unsupervised LDA topic models (discussed in Section B.8, OA). For consistency concerns, the research assistants are instructed to independently go through *all* sampled petitions and we compare their coding results afterwards. The inter-coder reliability is above 80% for both dimensions. When the RAs' coding decisions disagree, we read the petition and determine the proper classification by ourselves.

[12] We could not unambiguously determine the issue topic for about 4% of the petitions and petitioner's residence for about 19% of the petitions. These petitions are coded as missing.

[13] In doing so, we need to make two assumptions. The first is that that individuals who share similar grievances or social backgrounds use similar linguistic patterns, and the second is that the distribution

estimation procedure in Online Appendix B.5.

Table 2.1 displays the estimated proportions for issues and locations, all arranged in descending orders. For issues, the most frequently raised ones include (1) property transaction and management, (2) employment, (3) neighborhood environment, (4) land taking and house demolition, (5) education and health, (6) agriculture production, (7) labor disputes, and (8) social security. With the possible exception of (1),[14] the other seven of the top eight issues all seem to be life problems that are more commonly experienced by the lower class than by the middle or the upper class. Collectively, the seven issues account for about 50% of all petitions.

Table 2.1: Estimated Distributions for Issues and Petitioners' Backgrounds

Issue	Proportion	Location	Proportion
Property transaction/management	0.211	Urban	0.55
Employment	0.111	Rural	0.32
Neighborhood environment	0.099	Town/suburb	0.13
Land taking and house demolition	0.093		
Education/Healthcare	0.074		
Agricultural production	0.072		
Labor disputes	0.059		
Social security	0.052		
Business	0.046		
Traffic and transportation	0.046		
Public safety	0.044		
Infrastructure	0.041		
Hukou	0.026		
Corruption	0.024		
RMSE of 10-fold cross validation	0.023		0.057

Note: This table shows the estimated proportions for petition issues and petitioners' area of residence using semi-automated content analysis. For each model, we report the Root Mean Squared Error, which indicates the average deviation of estimated proportion from the true proportion in a 10-fold cross validation.

The composition of LLMB users can be seen even more clearly as we turn to the results

of word for each issue in the training set is representative of the population.

[14]The large proportion of property-related issues is likely to be due to their frequent occurrence in daily life. Our reading suggests that many of the complaints under this category are about malfunctions of facilities in one's apartments/residential compounds.

on location labels.[15] While the majority of the petitions still come from urban areas, about 45% of the petitions appear to be filed by users residing in rural and suburban areas, which are usually less developed economically.[16] As a benchmark, it is worth noting that as of 2016, rural residents still accounted for only about a quarter of active Internet users in China (CNNIC, 2016). Rural and lower-class users, in other words, are over-represented on the LLMB relative to their share in the Internet-using population.

2.5 The Policy Consequences of Online Participation

2.5.1 Data on Local Government Policy Priorities

To analyze the aggregate policy consequences of online participation, we made use of an original panel dataset of government policy priorities created from the text of local Government Work Reports (GWRs thereafter). GWRs are a form of official communication between the Chinese government and the legislative body at the same level. They are delivered once a year by the head of the government at the annual session of the People's Congress (PC) and have to be formally approved by PC deputies through an anonymous vote. As one of the most important policy documents that local administrations issue every year, GWRs provide comprehensive descriptions of governments' policy blueprints and highlight key socioeconomic targets for the following year. They have to be collectively read and edited by the party standing committees—the core leadership body in

[15]We determine petitioners' locations of residence based on the content of their petitions, which often mention where they live or even current address. Urban keywords include: residential complex, urban community, residential committee, urban district, inside a city. Rural: inside a village, village, village committee, in countryside, mountain village. Suburban: urban village, rural-urban fringe, township, central town, city outskirt, suburb.

[16]In the context of China, rural residents are in generally poorer than those living in urban areas. In 2008 (the launching year of LLMB), the average disposal income for rural residents was less than 1/3 of that of urban residents' (4761 yuan vs. 15781 yuan). See https://goo.gl/K3tfSp (in Chinese).

local governments—before they are sent out to the legislature for approval.

GWRs have a highly standardized structure: They often begin with a brief description of the overall national and local conditions, followed by a summary of governments' achievements in the past year; the bulk of the document is then devoted to laying out plans and directions for the next year. While the set of major policy areas that needs to be covered in a GWR is usually fixed (e.g., economy, public safety, culture, education, social welfare, etc), the relative amount of emphasis on each area is often subject to political discretion and can reflect local leaders' own policy visions. Newly published GWRs often receive close attention from media and government officials because they contain important information about leadership's policy preferences.

We collected the full text of over 4,400 government reports at both the city and provincial levels between 2000 and 2014,[17] and used a Latent Dirichlet Allocation (LDA) model (Blei et al., 2003) to uncover topics (i.e., cluster of words) from the text and estimate their relative proportions in each report. Compared to the traditional, dictionary-based coding methods, a distinct strength of the LDA algorithm is that it clusters words strictly according to their co-occurrence patterns, thus avoiding the arbitrariness and errors in hand coding. This is especially useful for analyzing complex policy documents such as the GWRs, for which commonly agreed upon coding rules do not yet exist. Moreover, since many words can be related to multiple policy areas, human coder may face difficulties in determining to which topic a particular word should be assigned. The LDA model provides a solution to this problem by allowing each word to be associated with multiple topics at different levels of strength.

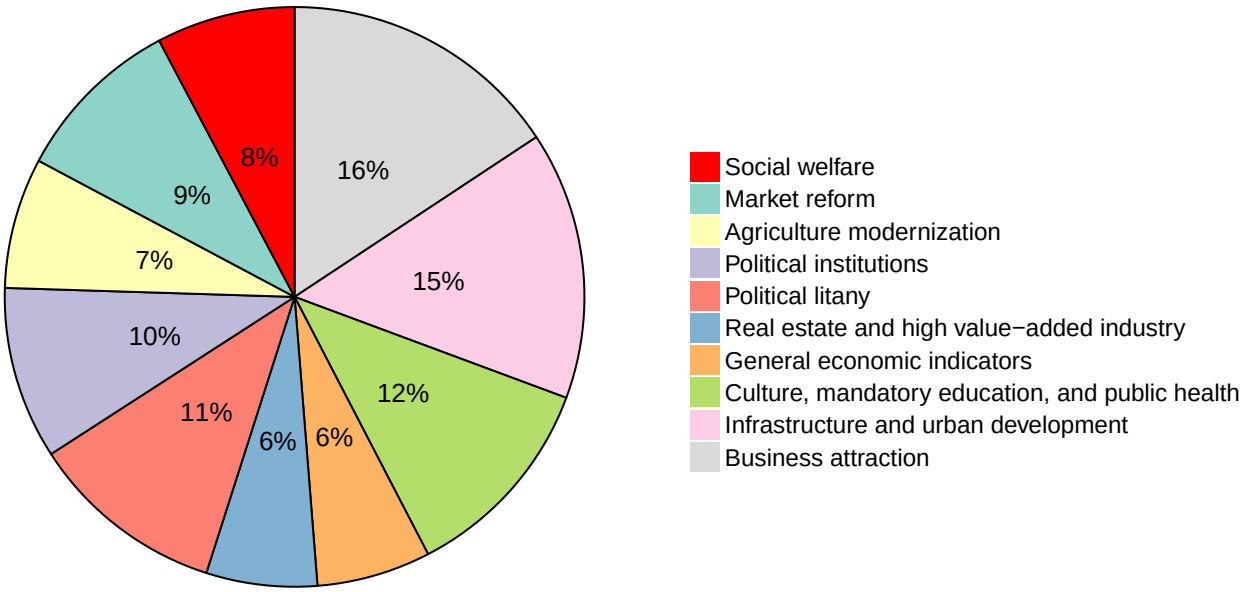
We estimated a 10-topic model as the baseline and a 20-topic model for robustness checks. Figure 2.1 plots the aggregate distribution of topic shares in the 10-topic model.[18]

[17]Full-text GWR data were collected from government websites and local yearbooks.

[18]The keywords with the highest association with each topic are detailed in Online Appendix B.6.4. We show that LDA is able to find highly meaningful policy areas. The output from the 20-topic model is available upon request.

In light of the previous findings on user background, we choose to focus on a topic for social welfare issues, which is a key concern for lower-class citizens.[19] In this model (as well as in the 20-topic model), we are able to detect a clear and highly coherent topic cluster on social welfare issues, which on average accounts for about 8% of a GWR. High frequency words under this topic include *social safety net*, *endowment insurance*, *medical insurance*, and *social insurance*. According to several post-estimation diagnostics that we performed, the social welfare topic is ranked as one of the highest-quality topics generated by both the 10- and 20-topic LDA models (see Online Appendix B.6 for detailed post-estimation diagnostics). A variance decomposition suggests that within-city variation accounts for about 63% of the total variation in this topic, whereas between-city variation accounts for the rest of the 37%.

Figure 2.1: Aggregated Distribution of GWR Topic Shares, 2000-2014



[19] In Table B2, we provide survey-based evidence on the difference in policy preferences for citizens of different socioeconomic backgrounds. Figure B6 presents results on the effect of LLMB petitions on *all* GWR policy topics.

2.5.2 Empirical Strategy

Our baseline is a fixed-effects model with the following specification:

$$\Delta\text{Welfare Topic Share}_{it+1} = \delta\text{Log Petitions}_{it} + \mathbf{X}_{it}\boldsymbol{\beta} + \phi_i + \tau_t + \epsilon_{it},$$

where i indexes cities and t indexes years.[20] The dependent variable, Δ *Welfare Topic Share*, is the incremental change in the proportion of welfare topic in a city's GWR from the previous year.[21] Because the delivery of the government work reports typically occur at the beginning of a year (January or February), we match each city-year spell with the change in GWR welfare topic in the following year.

The independent variable, *Log Petitions*$_{it}$ is the (logged) aggregate number of petitions about a city in a given year. This includes both petitions filed directly at the city's own message boards and those filed at the city's supervising provincial government regarding issues from a specific city.[22] Later, we also use a LDA model to detect distinct topic groups from the entire corpus of petitions and assign each individual petition to a specific topic.[23] This allows us to examine which subset of petitions has the strongest impact on welfare policy change. ϕ_i and τ_t represent city and year fixed-effects, respectively. With city fixed-effects, we difference out any variation that is city-specific but time-invariant, and exploit the within-city variation in petition volume and welfare policy. The main regression also includes a set of province-specific linear trends to account for influence

[20] Our main dataset covers 299 cities for the period of 2008 to 2013. Due to a high level of missingness in GWR data and relatively low usage of the LLMB, we drop three far-flung western provinces (Xinjiang, Tibet, and Qinghai) from the sample. Our results remain robust to including these provinces in the sample (Table B8).

[21] We choose to use change in welfare topic share as the dependent variable instead of its level to deal with the persistence and non-stationarity of welfare topic share.

[22] For the second type of petition, we identify their city origins based on posters' IP address and mentioning of city keywords in texts. When a petition mentions multiple city name keywords, we pick the unit that receives most mentioning as a petition's origin city.

[23] This procedure does not take input from the previous aggregate-level analysis. For each petition, the LDA model generates estimated proportions for all 30 topics. We assign each petition to the topic that takes up the highest proportion in that petition. For details, see Online Appendix B.8.

from time-variant, province-specific unobserved factors.[24] $\mathbf{X}_{it}$ is comprised of a rich set of time-variant controls, which are described below.

Socioeconomic Conditions

The most important potential confounder to our analysis is the intensity of public grievances expressed through other non-Internet channels.[25] Since dissatisfied citizens may concurrently use multiple means to make their grievances heard, changes in government policies may be attributable to their offline actions rather than online participation per se. Prior research has shown that collective protests are one of the most common means for lower-class citizens in China to express their discontent offline. Large-scale protests often receive a good deal of attention from the authority and can sometimes produce major shifts in government policies (Heurlin, 2016).

To address this, we make use of two of the best available datasets. The first one is the Collective Incidents Dataset, compiled by the Institute of Sociology at the Chinese Academy of Social Sciences (CASS). This dataset contains detailed information about major mass protests in China between 2007 and 2013, collected from both internal government documents and through extensive searches of domestic and overseas media reports. The second one, *China Strikes* (`www.chinastrikes.crowdmap.com`), is a crowd-sourced website that focuses specifically on labor-related unrest. For our analysis below, we combine all the unique events from both sources and calculate the frequency of protests for each city-year spell as a control variable.

In addition to protest frequency, we also include controls for a city's population *(Log Population*) and the size of employment (*Log Employment*), as unemployment is a com-

[24] An alternative specification is to use province-year dummies, which would consume more degrees of freedom in the estimation. Our main result is substantively the same when using this alternative specification.

[25] Although Chinese citizens do not enjoy the right to elect their government officials, prior research has shown that the regime does contain several non-electoral channels for citizens to make their voices heard (e.g., Shi, 1997; Tang, 2016).

mon source of popular grievances and a main driver for welfare expansion. Moreover, because welfare-related spending in China is covered primarily by local government budgets, it is likely to be affected by local economic and fiscal conditions. We thus include *Log GDP, GDP Growth Rate*, *Log Fiscal Revenue,* and *Log Fiscal Expenditure* to control for a locality's level and pace of development as well as fiscal capacity.

Leadership Characteristics

The second set of controls are about personal characteristics of local leaders. We include a number of demographic variables (for both the city secretary and the mayor), including age, tenure length, and political connection with the provincial leadership. Age and tenure length, in particular, have been found to be correlated with the career incentives of local leaders as well as their policy preferences. We also include the number of years a city leader has served in a given locality, as prior studies have shown that officials with longer local careers may be more attentive to local interests and spend more on social issues (Persson and Zhuravskaya, 2016).

2.5.3 Baseline Results

The main results are presented in Table 2.2. Column 1 presents the most parsimonious model with only two sets of fixed-effects and linear province trends as controls. We see that the (logged) total volume of online petitions is strongly and positively associated with governments' emphases on welfare policies. Specifically, the coefficient indicates that for an average city, a one standard deviation increase in online petition ($\sim$155 more petitions for a median city) is associated with a 0.48 percentage points, or about 17% of a standard deviation, increase in GWR welfare topic share. To put this figure in perspective: in a recent study of policy responsiveness in the United States, Caughey and Warshaw (2017) find that in the non-South, a one standard deviation change in mass liberalism is associated with about 3.7% of a standard deviation immediate change in social policy and

1.4% standard deviation change in economic policy. In another study, Miller (2015) finds that a one standard deviation loss in electoral vote in electoral autocracies is associated with about 30% of a standard deviation increase education and welfare spending. While such comparison should always be carried out with caution as these studies are based on very different measures of public sentiments, it nonetheless still provides some reassurance that the magnitude of our estimate is within a reasonable range.

Next, we decompose petitions into several more refined topic groups and examine their respective impacts on welfare policy change. We apply a 30-topic LDA model to the petition text and, based on the estimated topics, classify petitions by (1) whether they are from rural or urban areas, and (2) whether they involve issues directly related to one's pocketbook conditions. This gives rise to four distinct groups of petitions: rural petitions about pocketbook issues (RP), urban petitions about pocketbook issues (UP), rural petitions about non-pocketbook issues (RN), and urban petitions about non-pocketbook issues (UN). The details about the classification can be found in Online Appendix B.8.2. Type RP petitions are those related to rural low-income assistance, land-taking compensations, and financing of basic public goods. Common issues in type UP petitions include medical reimbursement, compensation for house demolition, wage disputes with employers, and unemployment benefits. Non-pocketbook issues include environmental degradation for rural areas and traffic control, public safety, education access, and property management for urban areas. Our expectation is that changes in welfare topics should be more strongly associated with types RP and UP petitions than with the other two types.

Columns 2 through 5 of Table 2.2 report the results from using the more refined petition groupings. Column 2 only makes the distinction between petitions concerning pocketbook and non-pocketbook issues, and Column 3 further distinguishes between petitions from rural and urban areas. Columns 4 and 5 incrementally add to the specification of Column 3 socioeconomic and leadership controls. The results largely confirm our expec-

Table 2.2: Main Results

	Δ Welfare Topic at $t+1$ (10-topic)				
	(1)	(2)	(3)	(4)	(5)
Log petitions	0.0048**				
	(0.0021)				
Log petitions (pocketbook)		0.0085***			
		(0.0029)			
Log petitions (non-pocketbook)		-0.0025			
		(0.0028)			
Log petitions (rural pocketbook)			0.0073***	0.0075***	0.0076***
			(0.0021)	(0.0022)	(0.0022)
Log petitions (urban pocketbook)			0.0019	0.0022	0.0021
			(0.0022)	(0.0023)	(0.0023)
Log petitions (rural non-pocketbook)			-0.0005	-0.0004	-0.0005
			(0.0016)	(0.0016)	(0.0016)
Log petitions (urban non-pocketbook)			-0.0023	-0.0026	-0.0026
			(0.0028)	(0.0028)	(0.0029)
City and year fixed-effects	✓	✓	✓	✓	✓
Province-specific trends	✓	✓	✓	✓	✓
Economic controls				✓	✓
Leadership controls					✓
R^2	0.02	0.03	0.03	0.04	0.04
Number of Cities	299	299	299	298	297
Observations	1656	1656	1656	1625	1624

Note: This table shows the results from OLS regressions. The dependent variables are incremental increase in the share of social welfare topic in government work reports. Socioeconomic controls include *Log Protest*, *Log Employment*, *Log Population*, *GDP Growth*, *Log Fiscal Revenue*, and *Log Fiscal Expenditure*. Leadership controls include (for both party secretary and mayor): *Age*, *Tenure*, and *Years of Local Service*, as well as an indicator for any city leader connected to the incumbent provincial secretary. Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

tation: Petitions concerning pocketbook issues appear to matter more for welfare policy change than those with non-pocketbook demands, and pocketbook petitions from rural areas appear to have the strongest association with changes in welfare topic share among the four types of petitions. For pocketbook petitions from urban areas, the coefficient is still positive but much smaller and less precisely estimated. One explanation for the weaker association may be that active participation by the middle class in urban areas has diverted some government attention from the lower class (as evidenced by the high frequency of property-related complaints).

2.5.4 Robustness and Endogeneity

We conduct several additional robustness checks on the main results. In the interest of space, we only provide a brief summary here and leave the details to the Online Appendix. We find that our main results are unchanged when we use welfare topic proportion generated by an alternative, 20-topic LDA model as the dependent variable (Table B7). Our results are also robust to including or excluding cities with special political status or distinct ethnic compositions (Tables B8). Moreover, we show that our results are not sensitive to removing cities with active local online petition platforms (Table B10).

We also take special care to address the issue of endogeneity. One major concern, for example, is that the volume of online petition may itself be a result of prior government welfare policies. To address this concern, we conduct several Granger-style tests. We find little evidence to support the reverse link: GWR welfare topic shares are not strongly correlated with petition volumes in subsequent years (Online Appendix B.14). Another concern is that both policies and online participation may be driven by some unobserved time-varying events, such as a shift in societal preference for social welfare. We address this issue through an instrumental variables (IV) analysis, which we detail in Online Appendix B.15. The IV results are largely consistent with the OLS ones.

2.5.5 Effects of Online Participation on Substantive Outcomes

The preceding analyses have demonstrated that the volume of LLMB petitions, especially those about rural, pocketbook issues, has a measurable impact on the emphasis of social welfare policies in government work reports. However, one may still question whether changes in policy rhetoric reflect actual changes in governing priorities. To address this issue, we examine the effect of online participation on more substantive policy outcomes. Our specific focus here is the Minimum Living Standard Guarantee Scheme (*dibao*), which is a major cash-based social assistance program targeted at the poor. Although this

program is not the only welfare program that the government implements, we choose to focus on it because it has the best available data and is unambiguously an issue that concerns the lower class. Researchers have shown that local governments typically have a lot of discretion in designing and implementing their own *dibao* programs (Solinger and Hu, 2012).[26] We thus expect that they will expand this program when pocketbook demands from lower-class groups become more salient.

We collect city-level data on the coverage of the rural *dibao* program from the website of the Ministry of Civil Affairs, and correlate them with type RP petitions (along with other types). Table 2.3 displays the regression results for three metrics: individual coverage, family coverage, and total spending. We can see that RP petitions have a positive and significant impact on the implementation of rural *dibao* across all these metrics. The instrumental variables analysis yields largely similar, if not stronger, results (Table B14). These results seem to suggest that online participation does have substantive redistributive consequences that go beyond policy rhetoric.

2.6 Concluding Remarks

As one of the most influential technological inventions of the 20th century, the Internet has been seen by many as having the potential to give citizens a louder voice in the political process. However, concrete evidence on whether and how Internet-based participation can affect government policies remains limited to date. This study addresses this gap by presenting new evidence on the policy consequences of online participation from China's largest electronic petition platform. Contrary to the prevailing view that in developing countries the Internet is still a luxury technology enjoyed mainly by the middle- and

[26]The *dibao* program is jointly funded by central and local governments, and local governments play an especially prominent role in raising money for the rural *dibao*, which is the focus of our empirical analysis here. Other responsibility of local governments include setting the maximum eligible income for local *dibao* applicants, approving *dibao* applications, disbursing the assistance, and monitoring the use of the fund (e.g., Bai and Gu, 2018).

Table 2.3: Change in Substantive Outcomes: Rural Minimum Living Standard Guarantee Scheme (*dibao*)

	Coverage (individual)	Coverage (family)	Total spending
	(1)	(2)	(3)
Log petitions (rural pocketbook)	0.0167**	0.0371***	0.0287***
	(0.0078)	(0.0127)	(0.0102)
Log petitions (urban pocketbook)	-0.0002	0.0008	-0.0170
	(0.0129)	(0.0150)	(0.0155)
Log petitions (rural non-pocketbook)	-0.0101	-0.0095	-0.0088
	(0.0062)	(0.0080)	(0.0095)
Log petitions (urban non-pocketbook)	-0.0108	-0.0202	-0.0106
	(0.0125)	(0.0167)	(0.0146)
City and year fixed-effects	✓	✓	✓
Number of Cities	301	301	301
Observations	1762	1762	1763

Note: The table reports results using several implementation outcomes of the rural *dibao* program as the dependent variable. The specification is based on Column 5 of Table 2.2.
Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

upper-class elites (Norris, 2001), our results suggest that when broadband infrastructure is made nearly universal, those from lower-class backgrounds will actively take advantage of new online channels to articulate their interests to the authorities, and that their participation can lead local governments to place higher priority on social welfare issues in both rhetoric and actual policy implementation.

Our study speaks to the broader literature on the relationship between the Internet and governance. Existing research has found that the Internet can help strengthen government accountability (Besley and Burgess, 2002) and reduce corruption (Bailard, 2009), but may also suppress voter turnout (Campante et al., 2013) and increase ideological polarization (Lelkes et al., 2015). On the issue of political inequality, more specifically, empirical findings from survey-based studies in advanced democracies (the U.S in particular) tend to support the view that the Internet will reinforce, rather than alleviate, the participation and influence gaps in the offline world (Norris, 2001; Schlozman et al., 2010). Our findings suggest that the conclusion drawn from the experience of developed countries may not necessarily apply to a developing context where the opportunities for

conventional forms of participation are more biased and limited. We provide evidence that supports the Internet's potential to mitigate participation inequality among citizens and show that bringing in new voices into politics can substantively change governments' policy priorities. While we certainly need to be cautious to not over-generalize findings from a particular platform, it is worth noting that some of the key features that made the platform effective, including operational independence, ease of access, and transparency, are not rare qualities in the cyberspace. As many other countries and regions have started to adopt similar electronic petition platforms, a promising direction of future research will be to investigate whether a similar equalizing impact also exists in other country settings.

This study also contributes to a nascent but rapidly growing literature on non-electoral responsiveness. Several recent studies have shown that in some non-Western political systems, and China in particular, inquiries and demands made by individual citizens receive replies from governments at a comparable rate as in electoral democracies (Chen and Kung, 2016; Distelhorst and Hou, 2017), and that individual politicians are generally attentive to citizen opinions (Meng et al., 2017; Truex, 2016). So far, however, there is still limited evidence on whether, in the absence of formal electoral accountability, the preferences and demands expressed by citizens will be systematically taken into account in government *policy making*—a more fundamental criterion of political responsiveness according to classical writings on this concept (Dahl, 1971; Miller and Stokes, 1963). Our findings suggest the existence of a form of responsiveness in China that matches with its canonical definition in the democratic context.

Finally, our analysis advances a more nuanced view on the relationship between the Internet and the durability of the Chinese regime. While previous studies tend to portray a largely confrontational picture, focusing on either the subversive capacity of the Internet or the regime's efforts to control and manipulate the cyberspace (King et al., 2013, 2017), our findings suggest that the interaction between the two is not always zero-sum: the democratic potential of the Internet may be harnessed by the government to fulfill

important governing functions. As such, the development of Internet may contribute to the regime's vitality by allowing it to improve the quality of governance without making more radical changes to its political institutions.

Chapter 3

Do Legal Rules Explain China's Economic Growth?

3.1 Introduction

The Chinese economy has been growing at a spectacular average annual rate of over 10% since the beginning of market reforms in the late 1970s. According to the World Bank, China's GDP in 2017 is 81.6 times as large as the country's GDP in 1978.[1] Paradoxically, the country is widely considered to have had poor property rights protection, high corrption, substantial costs of doing business, and a legal system heavily influenced by the will of the Party through much of the previous four decades (Allen et al., 2005). According to the World Bank's Doing Business indicators, China was ranked 78 in 2017, below Colombia, El Salvador, Indonesia, Ukraine, and many other less developed economies.

The *de jure* indicators of weak institutions in China appear to be at odds with the conventional wisdom that good domestic institutions are prerequisites for growth (La Porta et al., 2008; Acemoglu and Robinson, 2005; Acemoglu et al., 2019). That said, it is well-known that de jure institutions are often not enforced in many countries, due to selective implementation of laws and regulations (Hallward-Driemeier and Pritchett, 2015). China is a particularly interesting case given its decentralized power of economic policymaking at the local level and high powered incentives to develop the economy faced by local politicians, which gives rise to an economy that was often characterized as crony capitalism (Bai et al., Forthcoming), where firms with ties to the local government obtain resources

[1]The growth has slowed down since the global financial crisis. In a recent effort, Chen et al. (Forthcoming) estimates that the official Chinese GDP numbers overestimated actual GDP by 1.7 percentage points on average per year between 2008 and 2016. We bear this measurement error cautiously in mind.

at lower costs, and are able to circumvent inefficient regulations.

The purpose of this chapter is twofold. First, we characterize the evolution of China's legal landscape in the past four decades. We analyze a large corpus of 1.4 million legal documents issued by the Chinese government at central, provincial and city levels. Our corpus includes close to the universe of formal laws and around 1.35 million informal regulations.[2] Second, we explore whether the laws and regulations issued by provincial governments can explain variations in provincial economic growth in the reform era.

Since the transition is widely believed to be a triumph of markets over traditional central planning, we gauge the degree of market orientation in the legal documents using word embedding techniques to find words used in similar contexts as three "canonical" sets of words taken from, respectively, Karl Marx's *The Capital*, Gregory Mankiw's *Principles of Economics* and John Williamson's Washington Consensus. We find that the share of market-oriented language increased during the first half of the reform (1980-2000), but declined after around 2000.

We find that while there has been a rise in the frequency of market-oriented language in laws and regulations up to 2000, provinces' market reforms as manifested in the laws do not seem to contribute much to their GDP (per capita) and FDI growth. Specifically, market-oriented word shares explain very little of the variation in provincial growth on top of province and time fixed effects. Using a more flexible approach to characterize the legal documents (clustering) and select variables (LASSO), we also find that the documents have little predictive power for provincial growth. This suggests the possibility that businesses operate in significantly different environments from what is implied by legal

[2]Formal laws include national laws issued by the National People's Congress (*falv*), national regulations issued by the State Council (*xingzheng fagui*), national regulations issued by a department under the State Council (*bumen guizhang*), local decrees issued by a local People's Congress (*difangxing fagui*), and local rules issued by a local People's Government (*difang zhengfu guizhang*). Although not all of these documents are issued by People's Congress, the legislature, all of them have relatively sophisticated enactment procedures governed by the Legislation Law. Informal regulations are the rest of our corpus, most of which are called "normative documents" (*guifanxing wenjian*). The issuance of these documents is not governed by the Legislation Law, and is more flexible.

rules.

An extensive literature has studied reasons for China's economic miracle,[3] proposing factors ranging from the political system's emphasis on meritocracy (Li and Zhou, 2005) and relationship between central and local governments (Huang, 1996; Blanchard and Shleifer, 2001; Jin et al., 2005) to factional competition in the central government (Cai and Treisman, 2006). More recently, Bai et al. (Forthcoming) seeks to reconcile China's strong growth with weak de jure institutions by pointing to informal arrangements between local governments and firms and the incentives behind these arrangements.[4] Our study is a high-level yet systematic exploration of the relationship between legal rules and economic performance.

The rest of the chapter is organized as follows. Section 3.2 describes our corpus. Section 3.3 presents some descriptive statistics of lawmaking in China in the past four decades. Section 3.4 discusses our measurement of market orientation in the documents. Section 3.5 looks into whether the legal rules have predictive power for economic performance at province level. Section 3.6 concludes.

3.2 Data

We obtain our corpus of over 1.4 million Chinese legal documents from Chinalawinfo, a leading legal information technology company supervised by Peking University Law School. Chinalawinfo collected these documents conscientiously from government websites, government publications, and partner agencies.[5] The corpus contains close to the universe of formal laws (64,802) and around 1.35 million informal regulations. A few

[3]See comprehensive reviews by Malesky and London (2014) and Xu (2011).

[4]More broadly, the divergence between rules and actual conduct has been studied from both theoretical (Acemoglu et al., 2008) and empirical (Fisman and Wei, 2004; Carrillo et al., 2017) perspectives.

[5]These documents can be accessed with a subscription to Chinalawinfo's leading product PKULaw. However, bulk download is not possible through PKULaw. We got in touch with the company, and obtained their complete corpus for research purposes. To our knowledge, this is the first time that the entirety of PKULaw documents have been available to researchers.

recent papers have used small numbers of documents from Chinalawinfo to study specific Chinese policies (Chari et al., 2018; Fan, Forthcoming; Tian, 2019). We are not aware of any project that uses the entirety of these documents.

Although the informal regulations are more easily enacted, they are often quite important nonetheless. For instance, 99% of documents issued by central and local Development and Reform Commissions in our corpus, a key institution in China's economic policymaking, are "informal" regulations. Another example is that province-level minimum wages are stipulated in informal documents. For this reason, we find it important to study these documents as well even though we do not have the universe of them. We separate our analysis for formal and informal laws whenever convenient.

Figure 3.1 reports the number of documents in the corpus over time, broken down into formal and informal laws. The significant increase in the number of informal regulations in recent years is due to increased digitization and disclosure of government documents. A red line marks 2008, when the central government enacted a law mandating increased government transparency and information openness.

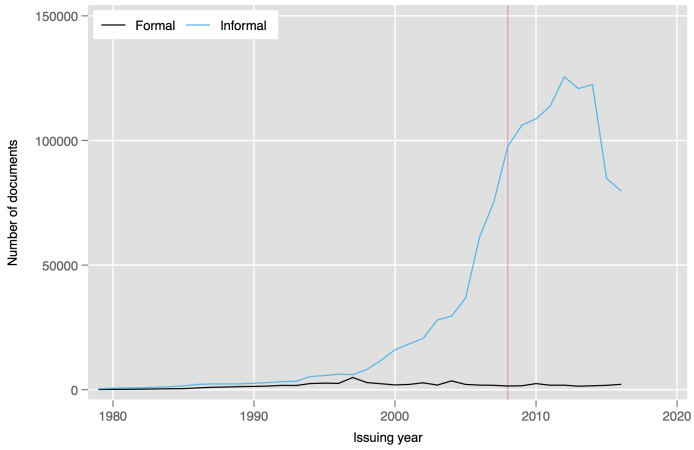


Figure 3.1: Number of Documents over Time

Notes: This figure shows the number of documents in the corpus over time, broken down into formal and informal laws.

Figure 3.3 provides a screenshot of the PKULaw interface. It reveals the standardized

template PKULaw uses to organize and report each policy document. Each document reported includes the following information: the government department that issued the document, the date of the issuance, the date on which the policy has been nullified and thus became ineffective.

Figure 3.3: Screenshot of a Sample Document

Notes: This figure shows a sample document on PKULaw.

PKULaw also assigned to each document a unique document ID and a policy category. Figure 3.5 shows the top 10 categories in terms of shares in our formal laws, out of a total of 104 categories. We also report the number of laws issued at each level of government. We see that areas such as real estate and environment protection are mostly regulated at the local level, while areas such as transportation and healthcare have large shares of central regulations.

3.3 Descriptive Statistics

We start our analysis by noting some patterns in lawmaking in the last four decades. We focus on formal laws in this section.

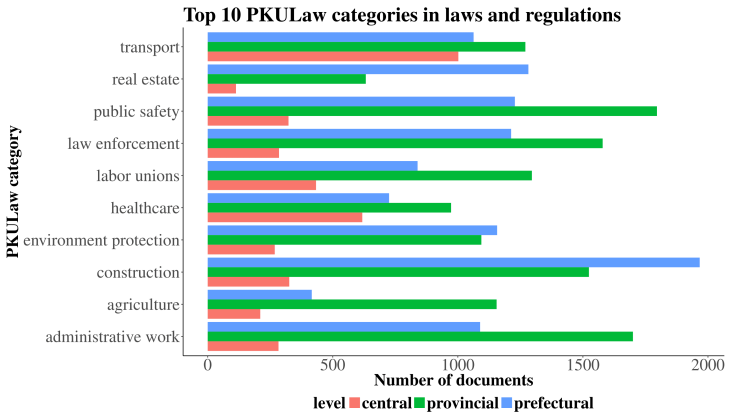


Figure 3.5: Top Formal Law Categories

Notes: This figure shows most prevalent formal law categories as well as the number of documents issued at each level of government.

3.3.1 Increasing lawmaking at local level

Figure 3.7 illustrates the number of formal laws issued each year and share of formal laws issued by local (province and prefecture) governments. As revealed by the red line, the share of local laws has been increasing rapidly from 1980 all the way to the mid 90s, and continues to increase gradually until the recent years. The black line shows that the number of formal laws enacted has been increasing rapidly from 1980 up to 1997, after which it dropped gradually.

3.3.2 Decreasing share of economic laws

Figure 3.9 shows the share of formal laws on economic matters issued by various levels of the Chinese government. Economics-related laws include all formal documents that belong to 70 3-digit PKULaw categories that appear to be obviously related to economic activities, ranging from contract laws, intellectual property rights, to labor laws and the governance of E-commerce. As revealed by the red line, the share of economics-related laws has been gradually declining over time.

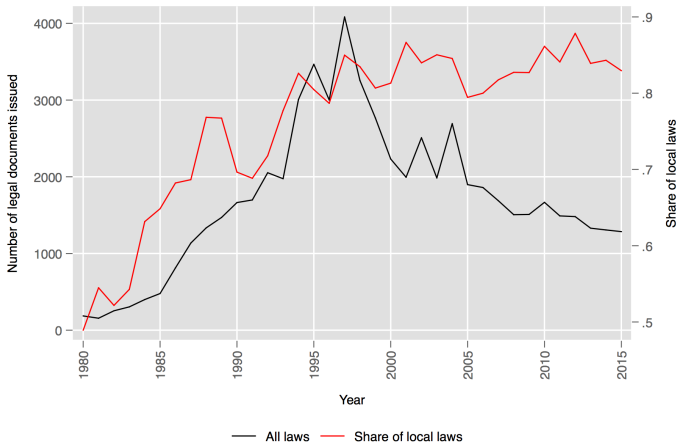


Figure 3.7: Share of Laws Issued by Local Governments

Notes: This figure shows number of formal laws enacted each year, as well as the share of formal laws issued by local (province and prefecture) governments.

Figure 3.9: Share of Laws on Economic Issues

Notes: This figure shows number of formal laws enacted each year, as well as the share of formal laws on economic issues, determined by PKULaw categories.

3.3.3 The phaseout of laws

We look at the lifespan of laws in this subsection.

Figure 3.11 shows the fraction of formal laws issued in year t (where $t < 2015$), that was still effective in 2015. The black line shows that less than 70% of the central laws issued before 2005 were still effective in 2015. Among the local laws issued before 2005,

their survival rates in 2015 were even lower. The blue line shows that less than 60% of the local laws enacted before 2005 were still effective in 2015. Laws that were enacted after 2005 had a substantially higher "survival" rate in 2015. Among laws made after 2005, local laws were slightly more likely to remain effective in 2015. These patterns show the substantial degree to which laws were updated.

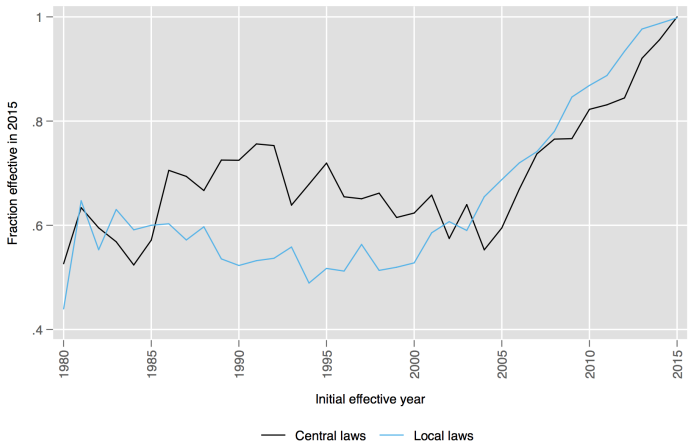


Figure 3.11: Share of Laws Effective in 2015

Notes: This figure shows the share of formal laws enacted each year that was still effective in 2015.

Another way to show that is to examine the effectiveness rate of laws n year since their issuance. Figure 3.13 shows that over 90% of laws still remained effective 2 years after their issuance. The fraction naturally dropped when we expand the time horizon. In particular, when we examine a law's effectiveness 10 years since its issuance, the survival rate is about 80% for laws issued before 2000, but dropped substantially to 70% for those that were issued between 2000 and 2007, indicating that laws passed in the 2000s are particularly short-lived.

In fact, the short life of laws made in the 2000s is largely caused by a large wave of modification and nullification of laws in 2010, a predetermined deadline for completing "the socialist legal system with Chinese characteristics".[6]

[6]See http://www.scio.gov.cn/zfbps/ndhf/2011/Document/1036756/1036756_1.htm

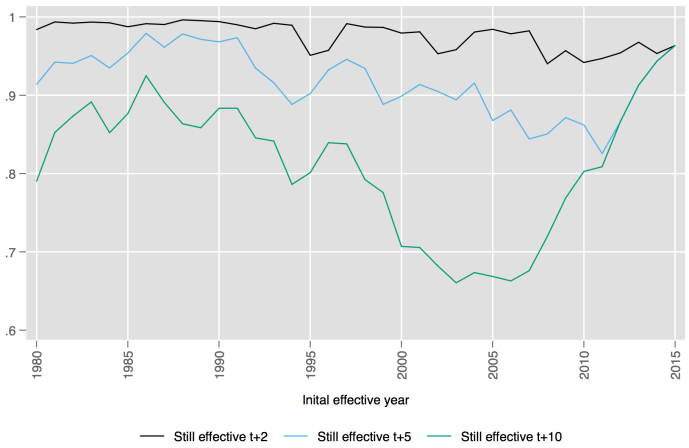


Figure 3.13: Share of Laws Still Effective n Years Since Issuance

Notes: This figure shows the share of formal laws enacted each year that was still effective in 2/5/10 years.

Yet another way to illustrate this nature of frequent turnover of Chinese laws is to examine the average effectiveness rate across all documents n years after their issuance, or a "survival curve". Two facts stand out from Figure 3.15. First, 10 years from the date of issuance, almost 20% of laws are no longer effective. Second, the average nullification and modification rate appears to be higher for economics-related laws, suggesting that the learning and turnover dynamics appear to be faster for economics-related policies.

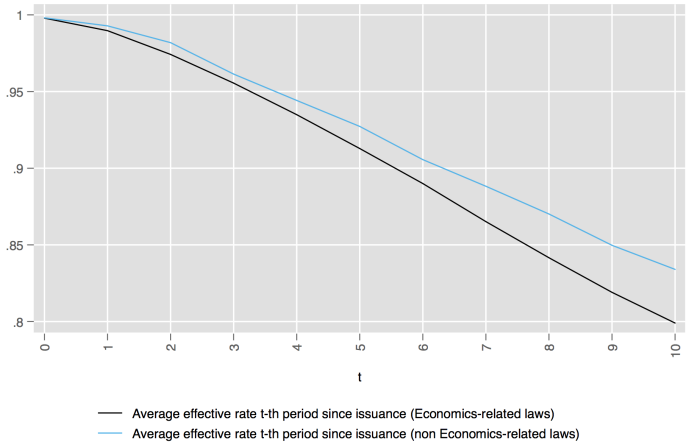


Figure 3.15: Survival Curves for Economic vs. Non-Economic Laws

Notes: This figure shows the share of formal laws that is still effective as a function of years since issuance, broken down into economic laws and non-economic laws.

3.3.4 The rise and fall of law areas

We examine the share of different areas of law in all formal laws, and calculate the change in share after 2000 from before 2000. Table 3.1 reports categories of laws that experienced largest percentage point changes. We see that post 2000, the areas experiencing largest decline in share are all economic areas, including public finance, labor unions, science and technology, etc, while the rising areas include legal institutions, administrative affairs, and environment protection. It appears that the focus of lawmaking after 2000 shifted from establishing basic economic institutions towards perfecting political institutions, which is also consistent with the pattern in Figure 3.9.

Table 3.1: Areas of Law Experiencing Largest Changes in Share

Increasing categories	% increase	Decreasing categories	% decrease
Legal institutions	5.8%	Public finance	2.7%
Administrative affairs	2.8%	Labor unions	2.2%
Construction	2.0%	Science & technology	1.5%
Environment protection	1.5%	Corporate finance	1.1%
Transportation	1.4%	Taxation	0.9%

Notes: This table reports categories of formal laws that experienced largest percentage point changes in share of all laws enacted after 2000 compared to before 2000.

3.4 Measuring Market Orientation in Policies

We now describe our approach of using Natural Language Processing techniques to measure market orientation of each document in our corpus.

Experts studying the Chinese economy have agreed that the spectacular growth of the Chinese economy is attributable to the government's market-oriented liberalization policies (Lardy, 2014; Naughton, 2018). While there are many anecdotes describing the country's economic reforms, we are not aware of a systematic analysis of the forty years of the reform process based on the government's legal documents. That is the goal of this

section.

3.4.1 A motivating example

Figure 3.17 shows the frequency of "commodity economy" and "market economy" mentioned per 1000 tokens (roughly meaning words) in the full corpus (both formal and informal laws). The frequency is essenetially zero in all documents before 1980, and exhibits interesting "peak and trough" patterns across time.

Specifically, the frequency of the word "commodity economy" shot up right after 1984, and remained high for the rest of the 1980s. In October 1984, leaders of the Chinese government introduced for the first time the notion of "commodity economy" into its economic lexicon, paving the way for further liberalization and market-oriented reforms. Previously, the word had been a ideological taboo due to its connections with capitalism. At the Third Plenary Session of CPC's 12th Central Committee in 1984, Party leaders determined that "the socialist planned economy is a commodity economy with plans on the basis of public ownership"[7], after a compromise had been reached between the progressive and conservative factions within higher echelons of the Party (Gewirtz, 2017). "Commodity economy" entered the official vocabulary ever since.

The frequency of the word "market economy" also exhibited interesting fluctuations. The frequency shot up right after 1992, when the CPC announced in the 14th National Congress of the Party to completely embrace the private sector as an key component of the economy, right after the famous Southern Tour by the architect of China's economic reforms Deng Xiaoping earlier that year. "Market economy" completely replaced "commodity economy" in legal documents as the country shedded more ideological baggage. The frequency of "market economy" shot up again around 2002, when in October that year the CPC announced the "preliminary completion of the market economy" in the 16th

[7]Source: Decisions on Economic System Reforms by CPC Central Committee. See http://cpc.people.com.cn/GB/64162/134902/8092122.html.

National Congress. Interestingly, the frequency of "market economy" dropped sharply afterwards, reflecting a shift in policy priorities.

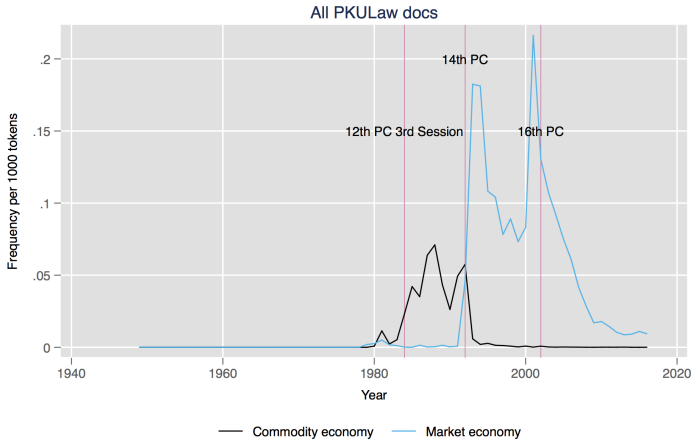


Figure 3.17: The Rise and Fall of "Commodity Economy" and "Market Economy"

Notes: This figure shows the frequency of "commodity economy" and "market economy" in all legal documents over time.

3.4.2 Pre-processing the corpus

Before conducting a text analysis, several steps need to be implemented. We pre-process the corpus through the following steps.

1. Remove all non-Chinese characters (e.g., "GDP", numbers).
2. Segment sentences into tokens using the "jieba" module (Chinese for "to stutter"). With the capacity to learn phrases from text, the module is able to recognize meaningful phrases such as "Three Represents".
3. Remove tokens containing just one character and a standard list of stop words (e.g, "some", "and", "because").
4. After these steps, we obtain over one million unique tokens, the majority of which provides little information for us to categorize documents, and increases computational burden. We thus filter out tokens whose frequency is particularly low or

appear in too many documents, with a cross-document sum of *tf-idf* lower than 100.[8]

The remaining tokens are sufficiently frequent, and do not show up in too many documents. Our pre-processed corpus is represented as a document-term matrix with 1414736 documents and 12693 tokens. With the document-term matrix constructed, we are ready to gauge the extent to which each Chinese policy document has moved towards markets from Marx.

3.4.3 Identifying market orientation in policy documents

While it is relatively easy to find a handful of words that have clear ideological implications (e.g., privatization, competition), handpicking keywords does not appear to be an objective approach. Therefore, we start with a set of words with relatively clear ideological orientation, and then use a technique called word embeddings to find words used in similar contexts as those clear-cut words in a data-driven fashion.

We obtain our initial anchor words from the following three sources:

1. Keywords in Karl Marx's classic *Das Kapital* discovered using the TextRank technique (Mihalcea and Tarau, 2004).[9]
2. "Neoclassical economics" words as appeared in the glossary of Gregory Mankiw's best-selling introductory economic textbooks for college students, *Principles of Economics*.
3. The ten policy areas suggested by the Washington Concensus as standard policy prescriptions for economic growth, first proposed by John Williamson in 1989.[10]

[8] See https://en.wikipedia.org/wiki/Tf%E2%80%93idf for the definition of tf-idf.

[9] The algorithm is similar to Google's PageRank algorithm for ranking websites, but applied to text instead. The basic idea is to find the most central words in a graph of text. Two words are connected in the graph if they appear within a window of each other. The importance of a word is determined recursively by examining the importance of words connected to it. We take the 50 most central words from *Das Kapital*.

[10] See Williamson (2004).

Table 3.2 reports the most frequent words (in the legal document corpus) in our three sets of "anchor words", as well as the share of each word in all occurrences of the set of words it belongs to.

Table 3.2: Anchor Words and Their Relative Shares

Panel A: Marx			
Word	Share	Word	Share
labor force	0.250	working class	0.005
worker	0.238	means of subsistence	0.003
ownership	0.150	industrial capital	0.002
productive forces	0.066	farm owner	0.002
means of production	0.057	equivalent	0.002
producer	0.056	capitalism	0.002
owner	0.054	amount of value	0.002
commodity circulation	0.019	usury	0.002
profit rate	0.017	bourgeoisie	0.002
mode of production	0.016	money capital	0.002
Panel B: Mankiw			
Word	Share	Word	Share
investment	0.236	currency	0.011
market	0.205	stock	0.010
social security	0.061	welfare	0.010
export	0.059	equilibrium	0.009
cost	0.039	market economy	0.009
consumption	0.039	diversification	0.009
import	0.036	equality	0.009
capital	0.035	principal	0.007
bond	0.033	screening	0.006
property rights	0.031	reserve	0.005
Panel C: Washington Consensus			
Word	Share	Word	Share
taxation	0.354	exchange rate	0.012
trade	0.269	fiscal deficit	0.001
property rights	0.233	fiscal expenditure	0.001
interest rate	0.105	privatization	0.0001
deregulation	0.025		

Notes: This table reports the most frequent words in our three sets of "anchor words", as well as the share of each word in all occurrences of the set of words it belongs to.

Word embeddings represent words as vectors to capture latent dimensions of meanings.

Mikolov et al. (2013a,b) propose two word embedding models, Continuous Bag-of-Words and Skip-gram, that explicitly target the relationship between words and their contexts (i.e., proximate words) in model training. Each word is represented by a vector. The cosine similarity between these vectors then measures the extent to which words are used in similar contexts.

We train the Continuous Bag-of-Words model on the entire PKULaw corpus. The model is a simple neural network that seeks to predict each word in the text from its surrounding context. Let $(w_1, w_2, ..., w_T)$ denote our corpus. The objective of the model is to maximize the log likelihood

$$\frac{1}{T}\sum_{t=1}^{T} \log\ p(w_t|w_{t-c}, w_{t-c+1}, ..., w_{t+c})$$

where $(w_{t-c}, w_{t-c+1}, ..., w_{t-1}, w_{t+1}, ..., w_{t+c})$ are the words surrounding w_t.

Abstracting from details of the neural network, the likelihood of a single word can be expressed as

$$p(w_t|w_{t-c}, w_{t-c+1}, ..., w_{t+c}) = \frac{\exp(v'^{T}_{w_t} \frac{1}{2c} \sum_j v_{w_{t+j}})}{\sum_{i=1}^{V} \exp(v'^{T}_{w_i} \frac{1}{2c} \sum_j v_{w_{t+j}})}$$

where v_w and v'_w are two vectors describing how word w is used in different contexts.

Intuitively, we see that if two words are often used in similar contexts (i.e., in proximity to a similar set of words), the algorithm is going to assign high cosine similarity to these two words. In training the model, we represent words as 400-dimensional vectors, and use a window size $c = 5$.

After we have the vector representation of words, we calculate the cosine similarity between each word in our corpus and each word in our three sets of anchor words.

As an illustration of the performance of the word embedding model, we list below the words identified as most similar (i.e., used in similar legal contexts) to "privatization".

Privatization (alternative Chinese word), demutualization, joint stock system, reor-

ganize, transform the system, corporation system, shareholding cooperative system, restructuring, transform the mechanism and build the system, asset restructuring, property rights system, state-owned enterprise, cooperative system, corporatization, change system, mixed system, shareholding cooperation, merger, debt-for-equity, state-owned

We see that the algorithm picks out words unique to the context of state-owned enterprise reforms in China. It would be very challenging for a human coder to consistently pick out these terms from over one million legal documents.

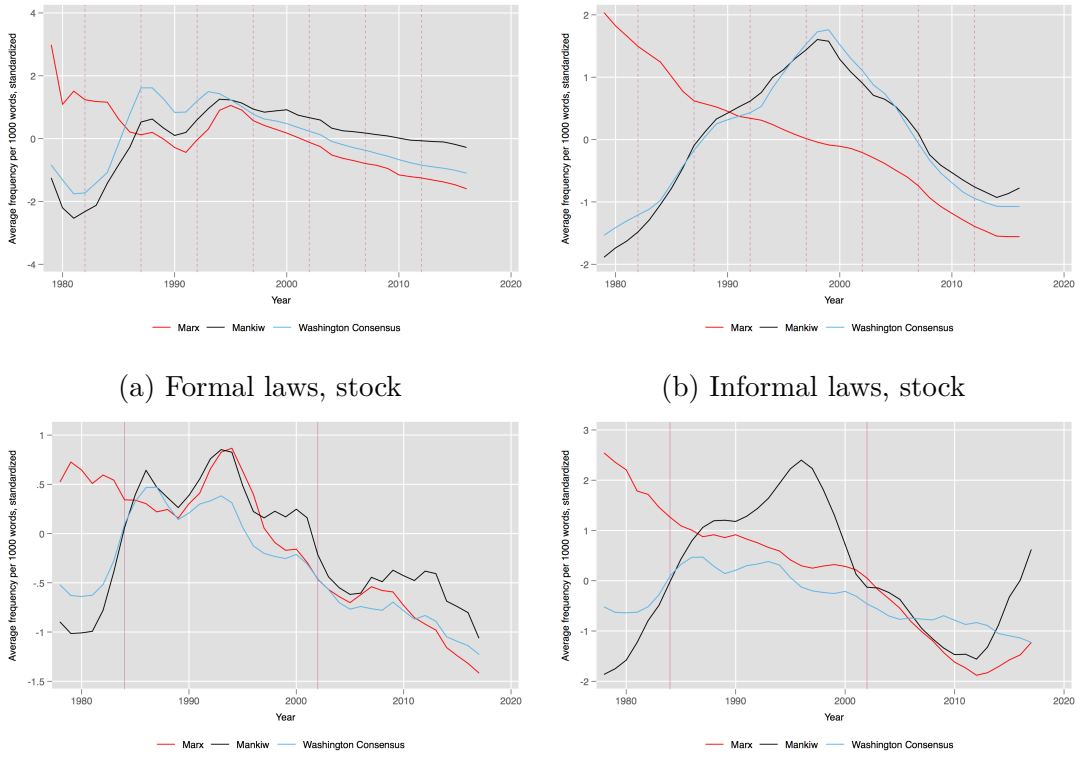
3.4.4 Results

3.4.4.1 Aggregate trends

Equipped with the word embeddings, we take all words above a similarity threshold to our anchor words, and calculate their share in each document. Since the shares depend on the number of anchor words as well as the similarity threshold, we do not find the level of shares interesting per se. Instead, we standardize the share to have mean 0 and standard deviation 1 across documents. We then plot their time series in Figure 3.19, where we separate formal and informal laws, and distinguish between the stock of laws, which is all laws enacted previously and that are effective in a given year, and the flow of new laws, which is the laws enacted in a given year. We choose similarity threshold 0.4, and plot five-year moving averages.

For both formal and informal laws, we see a steady drop of "Marxist" words over time, consistent with China's moving away from Marxist dogmas. For both economic (Mankiw) words and Washington Consensus words, we see a steady increase from the beginning of the reforms into the mid-90s, which then declined gradually. One possibility for this pattern is that by the end of the 90s, the CPC felt that much of the overdue economic reforms have been done, and it was time to focus on other social and political developments of the country. In fact, the peaks for both Mankiw words and Washington Consensus words fall roughly between 1984, the beginning of fully-fledged reforms and

2002, the "preliminary completion of socialist market economy", a pattern also seen in Figure 3.17. An exception to this pattern is an increase of Mankiw words in informal laws issued after 2012, the year when Xi Jinping took office.



(a) Formal laws, stock

(b) Informal laws, stock

(c) Formal laws, flow

(d) Informal laws, flow

Figure 3.19: Market Orientation in Legal Documents over Time

Notes: These figures report the share of words similar to the three sets of anchor words (Marx, Mankiw, Washington Consensus) in legal documents over time. Words with cosine similarity (to anchor words) higher than 0.4 are counted. Five-year moving averages are plotted. The shares are standardized to have mean 0 and standard deviation 1 across documents. Figures (a) and (b) plot for each year the documents that are effective in that year (i.e., the stock of laws). Figures (c) and (d) plot for each year new documents that are issued in that year (i.e., the flow of new laws). The red lines in figures (a) and (b) mark national Party Congress meetings that take place every five years. The two red lines in figures (c) and (d) mark 1984, the beginning of fully-fledged reforms and 2002, the "preliminary completion of socialist market economy".

Figure 3.21 illustrates the frequency of Washington Consensus words in laws enacted each year for coastal and inland provinces respectively.[11] For both regions, the hump

[11]Coastal provinces include Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Tianjin, Hebei, Shandong and Hainan.

shape is observed, with the former always having a higher frequency of Washington Consensus words, consistent with the common perception that coastal provinces have more liberal and outward-oriented policies. It is also interesting to note that Washington Consensus words peaked earlier in formal laws between the mid-80s and the mid-90s compared to in informal laws, where they peaked in the 90s.

(a) Formal laws (b) Informal laws

Figure 3.21: Washington Consensus over Time, Coastal vs. Non-Coastal Provinces

Notes: These figures report the share of words similar to Washington Consensus anchor words in legal documents enacted each year, broken down into coastal and non-coastal provinces. Words with cosine similarity (to anchor words) higher than 0.4 are counted. Five-year moving averages are plotted. The shares are standardized to have mean 0 and standard deviation 1 across documents.

The Appendix contains aggregate trends for each of the ten Washington Consensus components, where we see that with the notable exception of "deregulation", all policy areas suggested by the Washington Consensus experienced a hump shape in the Chinese legal documents.

3.4.4.2 Regression evidence

The aggregate trends we observe could be simply due to changes in the composition of areas of law. Here we look at whether these dynamics are present even within fine-grained areas of law.

We estimate the following specification

$$Z(M_{ict}) = \alpha_{\text{prov}(i)} + \lambda_c + \beta_t + \epsilon_{ict} \tag{3.1}$$

where $Z(M_{ict})$ is the z-score of the share of the words similiar to one of the three anchors in document i that belonged to PKULaw subcategory c and is issued in period t. α are province fixed effects. λ are subcategory fixed effects. These are very fine-grained areas of law. There are a total of 1159 subcategories, with a typical subcategory being "land use of foreign enterprises". β are 5-year period fixed effects corresponding to the 5-year plan time frames. We cluster standard errors at the level of rough categories.

Figure 3.23 reports estimates of the time fixed effects. We see that for both formal and informal documents, there is a downward trend in the Marxist language over the entire sample period. The downward trend is significantly more pronounced for the informal laws. Despite the fact that we see a hump shape in the aggregate trend of Mankiw words, within fine areas of law, Mankiw words become increasingly common over time. For Washington Consensus words, the hump shape is present even within areas of law, and we again see the peak appearing later in the informal laws than in the formal laws.

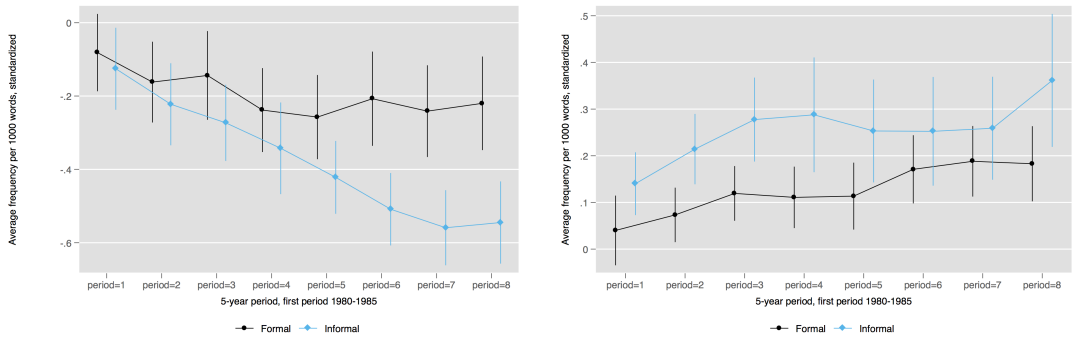
3.5 Do Legal Rules Explain Economic Growth?

Our goal now is to study whether the legal rules can explain differential economic growth across provinces.

3.5.1 Regressions with word shares

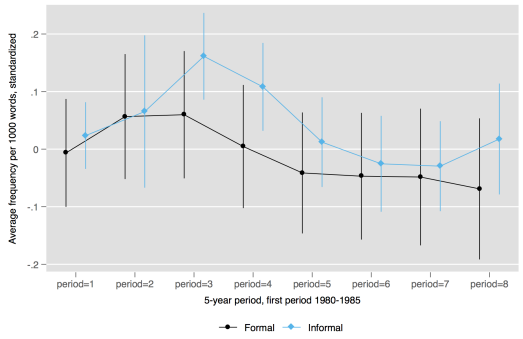
We estimate the following specification

$$\Delta y_{pt} = \alpha_p + \lambda_t + \beta \text{word share}_{\text{pt}} + \gamma' X_{pt} + \epsilon_{pt}, \tag{3.2}$$



(a) Marx

(b) Mankiw



(c) Washington Consensus

Figure 3.23: Trends of Market Orientation within Fine Areas of Law

Notes: These figures report estimates of time fixed effects in Equation 3.1. The five-year periods correspond to the five-year plans, with the first period being 1980-1985.

where Δy_{pt} is the growth rate of GDP per capital in province p over 5-year period t. α is a province fixed effect. λ is a 5-year period fixed effect. word share$_{\text{pt}}$ is the average frequency share of one of the three sets of anchor-related words in the "stock" of regulations ever issued by the province that still remained effective by period t. In addition to the fixed effects, we also include control variables including log number of regulations and log average length of regulations.

Table 3.3 shows the results of estimating Equation 3.2. As column (1) shows, simply adding province fixed effects can only account for about 4.2% of the variation in gdp per capita growth across provinces and time. When year fixed effects are added in column (2), the R-squared shoots up to 78%. Notice that year fixed effects capture all variation in the central government's national laws across time. The significant increase in R-squared

suggests that national laws and other macroeconomic factors still play an important role, compared to local regulations, in driving provincial economic growth. In column (3), we add the (log) share of Mankiw-related words in formal and informal documents separately as regressors, in addition to the year and province fixed effects. We find that the intensity of the Marxist language in both sets of documents does not explain provincial GDP outcomes. In columns (4) and (5), we repeat the exrecise of columns (3) by replacing the two regressors of Marxist word shares by Mankiw words shares and WC word shares, respectively. None of them appear to predict provincial GDP per capita. When all of the six regressors of anchor-related word shares are included as regressors, none besides the WC-related word shares in informal documents of the province are positively correlated with provincial economic growth.

One may think that there are many channels through which market-oriented policies affect the structure rather than the level of economic outcomes. Given the drastic increase in trade flows and foreign direct investment (FDI) into China since its government's trade and FDI liberalization policies in the mid 90s, we may expect to see a stronger correlation between the word count shares of market-oriented words in policy documents and performance in provincial trade or FDI activities. We thus replace GDP per capita growth in Table 3.3 by the mean FDI/GDP ratio over a 5-year period. Table 3.4 shows the estimation results. Column (1) shows that province fixed effects explain significant variation in FDI outcomes, compared with economic growth as reported in Table 3.3. Specifically, the R-squared is 54% just with province fixed effects included as a regressor, suggesting that province time-invariant characteristics (e.g., coastal provinces) explain much of the observed variation in FDI across provinces. In column (2), when year fixed effects are included as regressors additionally, the R-squared increases to 66%. In column (3), when Marx-related word share in formal and informal documents are separately added as regressors, the R-squared increases to 70%, with the (log) share of Marx-related words in informal documents issued by the province being negatively and significantly

Table 3.3: Do Legal Rules Explain GDP Growth?

	DV: provincial GDP per capita growth over 5-year period					
	(1)	(2)	(3)	(4)	(5)	(6)
log Marx, formal			0.007			0.014
			(0.028)			(0.062)
log Marx, informal			0.041			0.083
			(0.042)			(0.051)
log Mankiw, formal				0.025		0.073
				(0.055)		(0.088)
log Mankiw, informal				-0.019		-0.114
				(0.059)		(0.072)
log WC, formal					0.017	-0.023
					(0.029)	(0.037)
log WC, informal					0.028	0.054**
					(0.022)	(0.025)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	No	Yes	Yes	Yes	Yes	Yes
Observations	217	217	211	211	206	206
R-Squared	0.042	0.784	0.792	0.790	0.796	0.803
Adjusted R-Squared	-0.113	0.741	0.740	0.738	0.744	0.746

Notes: This table reports estimates of Equation 3.2. Standard errors are clustered at province level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

correlated with the 5-year average FDI/GDP ratio.

We repeat the same estimations of column (3), but with the two Marx-related word shares replaced by the corresponding Mankiw-related shares in column (4), and by the corresponding WC-related shares in column (5). We find that while the intensity of Mankiw language does not seem to matter for FDI, the WC word shares in informal documents of the province are positively correlated with provincial FDI.

In sum, based on the small improvement in R-squared when the market-oriented word shares are included as regressors for both sets of regressions, we find no evidence that market oriented policies can explain province-level macroeconomic outcomes.

Table 3.4: Do Legal Rules Explain FDI?

	DV: mean FDI-to-GDP ratio over 5-year period					
	(1)	(2)	(3)	(4)	(5)	(6)
log Marx, formal			-0.000			-0.002
			(0.001)			(0.002)
log Marx, informal			-0.002**			-0.001
			(0.001)			(0.001)
log Mankiw, formal				-0.001		0.002
				(0.001)		(0.004)
log Mankiw, informal				0.000		-0.000
				(0.001)		(0.001)
log WC, formal					-0.002**	-0.001
					(0.001)	(0.001)
log WC, informal					0.001**	0.001*
					(0.001)	(0.001)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	No	Yes	Yes	Yes	Yes	Yes
Observations	202	202	197	197	192	192
R-Squared	0.535	0.660	0.700	0.689	0.710	0.717
Adjusted R-Squared	0.457	0.588	0.621	0.607	0.631	0.630

Notes: This table reports estimates of Equation 3.2. Standard errors are clustered at province level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.5.2 Accounting for policies more flexibly

While the previous regressions with word shares suggest the limited predictive power of the legal rules for provincial economic outcomes, it is possible that the word shares alone do not fully capture changes in laws and regulations. Here we characterize the legal rules in a very flexible way, and examine whether they can predict GDP growth and FDI at province level.

The basic challenge we face here is that the number of legal documents is much larger than the number of observations (province-time). We tackle this challenge in two steps. First, we group together closely related documents to reduce the number of "policies" we consider. Second, we use LASSO to select variables on the basis of out-of-sample

predictive power.

3.5.2.1 Identifying policies through document clustering

We cluster the documents using DBSCAN (Ester et al., 1996), a widely used clustering algorithm.[12] The notion of distance between documents we use is cosine similarity in *tf-idf* vectors.

We call the document that is first issued in a cluster of documents a "bellwether" law. Table 3.5 presents provinces whose "bellwether" laws as a share of all formal laws enacted by the province is the highest. This measures the "innovativeness" in lawmaking of different provinces. We see that the central government is by far the most innovative legislative entity, with 8.9% of its formal laws being bellwether laws. The central government is followed by Shanghai, Beijing and Guangdong, all of which are commonly perceived as pioneers in reforms.[13]

3.5.2.2 Predictive power of policy clusters

Equipped with the clusters, we use LASSO to explore which, if any, clusters can predict provincial growth. Let c be a generic cluster collected in the set $\mathscr{C}$. We focus on laws that are categorized as economics-related, and include both formal and informal laws in the analysis.

[12]The algorithm does clustering by growing clusters incrementally from "core points". Once a point is included in a cluster and it is a core point, all points that are within a given radius of the point are subsequently included. Two intuitive parameters are required: the minimum number of points in a point's neighborhood for that point to be a core point (which is also the minimum number of points in a cluster), n, and the radius with which clusters grow, ϵ. Unlike algorithms such as k-means, the number of clusters, which we have no prior knowledge about, does not need to be specified.

ϵ controls the granularity of policies - a larger ϵ produces coarser policy clusters. Choosing a smaller ϵ will generally produce more documents that are not included in any cluster (outliers), but will result in more fine-grained clusters. We choose $\epsilon = 0.3$. Meanwhile, we choose n to be the small number of 3, as we want to capture those policies that never successfully spread.

[13]We look at formal laws in this exercise, for which we have comprehensive sample coverage. Therefore, the ranking is not caused simply by the fact that documents issued by the central government, Shanghai, Beijing, and Guangdong are more likely to be collected by PKULaw.

Table 3.5: Provinces with Highest Shares of Bellwether Laws

Issuing entity	Share of laws bellwether
Central government	8.87%
Shanghai	5.75%
Beijing	5.21%
Guangdong	3.29%
Heilongjiang	3.04%
Jilin	2.80%
Tianjin	2.66%
Hainan	2.43%
Fujian	2.22%
Sichuan	2.12%

Notes: This table reports provinces where the shares of "bellwether" laws in all formal laws issued by the province are highest. A bellwether law is defined as the earliest law in a cluster of closely related laws nationwide. Laws are clustered using the DBSCAN algorithm according to similarity in full text, with parameters minPts = 3, $\epsilon = 0.3$.

Specifically, we estimate the following relationship.

$$y_{pt} = \gamma_p + \lambda_t + \sum_{c \in \mathscr{C}} \beta_c f_{cpt} + \epsilon_{pt}$$

where γ_p and λ_t are province and period fixed effects. $f_{cpt} = 1$ if policy cluster c is effective in province p throughout period t.

We use LASSO to estimate this equation, forcing the fixed effects to be in the equation by setting their panelties at 0. We choose the panelty coefficient λ by minimizing out-of-sample MSE in a cross-validation.

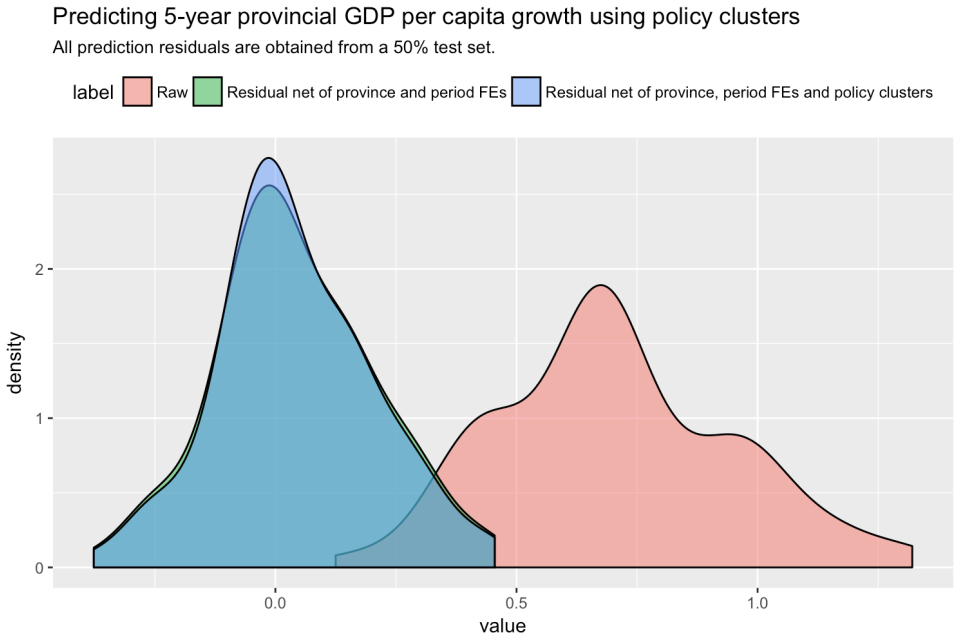
To evaluate the predictive power of the legal rules, we perform the following exercise. We plot side by side three distributions: the distribution of the outcome, the distribution of residuals of the outcome net of fixed effects, and the distribution of residuals of the outcome net of fixed effects and estimated effects of the policy clusters. In the latter two cases, the residuals are out of sample, i.e., the model is first estimated on a training set, and the residuals are then obtained from a test set.

Figure 3.25 presents results for both GDP per capita growth and FDI/GDP ratio.

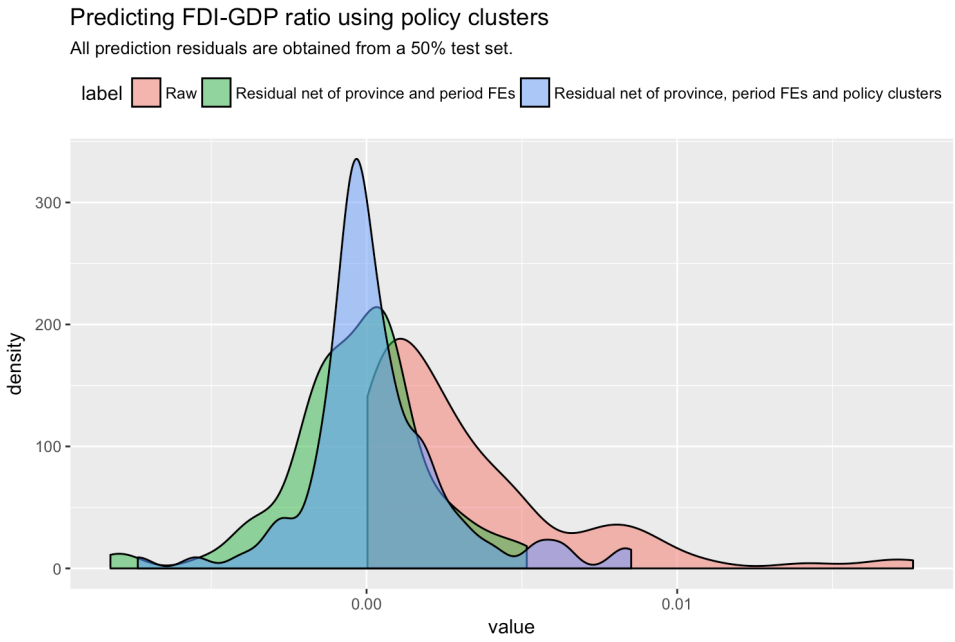
We see that even this very flexible representation of the legal rules can predict very little of the extra variation in provincial economic outcomes on top of province and period fixed effects. In the case of GDP growth, province and period fixed effects reduce the variance in GDP per capita growth by 53.7%. The policy clusters further reduce variance by just 2.33%. For FDI/GDP ratio, province and period fixed effects reduce the variance in FDI/GDP ratio by 56.7%, and adding policy clusters actually increases variance by 2.78% due to overfitting.

3.6 Conclusion

In this chapter, we find that the Chinese government actively introduced pro-market institutions from the mid 1980s to around 2000, which then slowed down after 2000. However, the market orientation of regulations only explains a small fraction of the provincial variation in GDP per capita growth and FDI. A richer representation of the documents also exhibits small predictive power. Taken together, this suggests the importance of studying the informal arrangements between market participants and government officials in more detail, along the lines of Hallward-Driemeier and Pritchett (2015) and Bai et al. (Forthcoming).



(a) GDP per capita growth



(b) FDI/GDP ratio

Figure 3.25: Predictive Power of Legal Rules

Notes: These figures plot side by side three distributions: the distribution of the outcome (GDP per capita growth or FDI/GDP ratio), the distribution of residuals of the outcome net of fixed effects, and the distribution of residuals of the outcome net of fixed effects and estimated effects of the policy clusters. In the latter two cases, the residuals are out of sample, i.e., the model is first estimated on a training set, and the residuals are then obtained from a test set.

Bibliography

Acemoglu, D., G. Egorov, and K. Sonin (2008): "Coalition Formation in Non-Democracies," *The Review of Economic Studies*, 75, 987–1009.

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2019): "Democracy does cause growth," *Journal of Political Economy*, 127, 47–100.

Acemoglu, D. and J. A. Robinson (2005): *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Alesina, A. and E. L. Ferrara (2005): "Preferences for Redistribution in the Land of Opportunities," *Journal of Public Economics*, 89, 897 – 931.

Allen, F., J. Qian, and M. Qian (2005): "Law, finance, and economic growth in China," *Journal of Financial Economics*, 77, 57–116.

Almond, D., Y. Chen, M. Greenstone, and H. Li (2009): "Winter heating or clean air? Unintended impacts of China's Huai River policy," *American Economic Review: Papers & Proceedings*, 99, 184–190.

Arzaghi, M. and J. V. Henderson (2008): "Networking off Madison Avenue," *Quarterly Journal of Economics*, 75, 1011–1038.

Asami, Y., X. Gao, and W. Katsumata (2006): "Evaluating land-use restrictions concerning the floor area ratio of lots," *Environment and Planning C: Government and Policy*, 24, 515–532.

Audretsch, D. and M. Feldman (1996): "R&D spillovers and the geography of innovation and production," *American Economic Review*, 86, 630–640.

Bai, C. and X. Gu (2018): "Horizontal Inequality in China's Social Safety Net," *Chinese Public Administration*, 1, 109–115, in Chinese.

Bai, C.-E., C.-T. Hsieh, and Z. M. Song (Forthcoming): "Special deals with Chinese characteristics," *NBER Macroeconomics Annual*.

Bailard, C. S. (2009): "Mobile Phone Diffusion and Corruption in Africa," *Political Communication*, 26, 333–353.

Barjamovic, G., T. Chaney, K. Coşar, and A. Hortaçsu (2017): "Trade, merchants, and the Lost Cities of the Bronze Age," Tech. rep.

Barr, J. and J. P. Cohen (2014): "The floor area ratio gradient: New York City, 1890-2009," *Regional Science and Urban Economics*, 48, 110 – 119.

Bertaud, A. (2018): *Order Without Design: How Markets Shape Cities*, MIT Press.

Bertaud, A. and J. K. Brueckner (2004): "Analyzing building height restrictions - predicted impacts, welfare costs, and a case study of Bangalore, India," Tech. rep., The World Bank.

Besley, T. and R. Burgess (2002): "The Political Economy of Government Responsiveness: Theory and Evidence from India," *The Quarterly Journal of Economics*, 117, 1415–1451.

Bimber, B. (1998): "The Internet and Political Transformation: Populism, Community, and Accelerated Pluralism," *Polity*, 31, 133–160.

Blanchard, O. and A. Shleifer (2001): "Federalism with and without Political Centralization: China Versus Russia," *IMF Staff Papers*, 48, 171–179.

Bleakley, H. and J. Lin (2012): "Portage and Path Dependence," *The Quarterly Journal of Economics*, 127, 587–644.

Blei, D. M. (2012): "Probabilitic Topic Models," *Communications of the ACM*, 55, 77–84.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Brockett, D., D. Fridley, J. Lin, and J. Lin (2002): "A tale of five cities: the China residential energy consumption survey," *ACEEE Summer Study on Building Energy Efficiency*, 29–40.

Brueckner, J. K., S. Fu, Y. Gu, and J. Zhang (2017): "Measuring the stringency of land use regulation: The case of China's building height limits," *Review of Economics and Statistics*, 99, 663–677.

Brueckner, J. K. and S. V. Lall (2015): "Cities in developing countries: fueled by rural–urban migration, lacking in tenure security, and short of affordable housing," in *Handbook of regional and urban economics*, Elsevier, vol. 5, 1399–1455.

Cai, H. and D. Treisman (2006): "Did Government Decentralization Cause China's Economic Miracle?" *World Politics*, 58, 505–535.

Campante, F. R. and Q.-A. Do (2014): "Isolated capital cities, accountability, and corruption: Evidence from US states," *American Economic Review*, 104, 2456–2481.

Campante, F. R., Q.-A. Do, and B. Guimaraes (2014): "Capital cities, conflict, and misgovernance: Theory and evidence," Sciences Po Economics Discussion Papers 2014-13, Sciences Po Departement of Economics.

Campante, F. R., R. Durante, and F. Sobbrio (2013): "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation," Working Paper 19029, National Bureau of Economic Research.

Carrillo, P., D. Pomeranz, and M. Singhal (2017): "Dodging the taxman: firm misreporting and limits to tax enforcement," *American Economic Journal: Applied Economics*, 9, 144–164.

Castells, M. (1996): *The Rise of The Network Society: The Information Age: Economy, Society and Culture*, Wiley.

Caughey, D. and C. Warshaw (2017): "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014," *American Political Science Review*, 112, 249–266.

Chadwick, A. and C. May (2003): "Interaction between States and Citizens in the Age of the Internet: "e-Government" in the United States, Britain, and the European Union," *Governance*, 16, 271–300.

Chari, A., E. Liu, S.-Y. Wang, and Y. Wang (2018): "Property rights, land misallocation and agricultural efficiency in China," Tech. rep.

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (2017): "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States," *Journal of Urban Economics*, 98, 17–49.

Chen, T. and J.-S. Kung (2016): "Do land revenue windfalls create a political resource curse? Evidence from China," *Journal of Development Economics*, 123, 86 – 106.

Chen, W., X. Chen, C.-T. Hsieh, and Z. M. Song (Forthcoming): "A forensic examination of China's national accounts," *Brookings Papers on Economic Activities*.

CNNIC (2016): "2015 Report on Rural Internet Development," .

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux (2010): "Estimating agglomeration economies with history, geology, and worker effects," in *Agglomeration Economies*, ed. by E. L. Glaeser, Chicago University Press, Chicago.

Cui, H., X. Li, and W. Zhang (2005): "China's inter-census survey in 2005," *Beijing: National Bureau of Statistics of China*.

Dahl, R. A. (1971): *Polyarchy: Participation and Opposition*, New Haven, CT: Yale University Press.

Davis, D. R. and D. E. Weinstein (2002): "Bones, bombs, and break points: the geography of economic activity," *American Economic Review*, 92, 1269–1289.

Dingel, J. I., N. A. Miscio, and D. R. Davis (2018): "Cities, lights, and skills in developing economies," .

Distelhorst, G. and Y. Hou (2017): "Constituency Service under Nondemocratic Rule: Evidence from China," vol. 79, 1024–1040.

Dobbs, R., Y. Chen, G. Orr, J. Manyika, M. Chui, and E. Chang (2013): "China's E-tail Revolution," Tech. rep., McKinsey Global Institute.

Donaldson, D. and R. Hornbeck (2016): "Railroads and American economic growth: A 'market access' approach," *The Quarterly Journal of Economics*, 131, 799–858.

Dutton, W. H. (2009): "The Fifth Estate Emerging through the Network of Networks," *Prometheus*, 27, 1–15.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996): "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, OR, AAAI Press, vol. 96, 226–231.

Fan, J. (Forthcoming): "Internal geography, labor mobility, and the distributional impacts of trade," *American Economic Journal: Macroeconomics*.

Fang, H., Q. Gu, W. Xiong, and L.-A. Zhou (2016): "Demystifying the Chinese housing boom," *NBER macroeconomics annual*, 30, 105–166.

Fick, S. E. and R. J. Hijmans (2017): "WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas," *International Journal of Climatology*, 37, 4302–4315.

Fischer, G., F. Nachtergaele, S. Prieler, H. van Velthuizen, L. Verelst, and D. Wiberg (2008): "Global agro-ecological zones assessment for agriculture (GAEZ 2008)," *IIASA, Laxenburg, Austria and FAO, Rome, Italy*.

Fisman, R. and S.-J. Wei (2004): "Tax rates and tax evasion: evidence from "missing imports" in China," *Journal of political Economy*, 112, 471–496.

Gabaix, X. and R. Ibragimov (2011): "Rank- 1/2: a simple way to improve the OLS estimation of tail exponents," *Journal of Business & Economic Statistics*, 29, 24–39.

Gaughan, A. E., F. R. Stevens, Z. Huang, J. J. Nieves, A. Sorichetta, S. Lai, X. Ye, C. Linard, G. M. Hornby, S. I. Hay, H. Yu, and A. J. Tatem (2016): "Spatiotemporal patterns of population in mainland China, 1990 to 2010," *Scientific Data*, 3, 160005.

Gewirtz, J. (2017): *Unlikely partners*, Harvard University Press.

Glaeser, E. L. and J. D. Gottlieb (2009): "The wealth of cities: agglomeration economies and spatial equilibrium in the United States," *Journal of Economic Literature*, 47, 983–1028.

Glaeser, E. L., J. Gyourko, and R. Saks (2005): "Why is Manhattan so expensive? Regulation and the rise in housing prices," *The Journal of Law and Economics*, 48, 331–369.

Glaeser, E. L. and D. C. Mare (2001): "Cities and skills," *Journal of labor economics*, 19, 316–342.

Goodman, D. S. G. (2014): *Class in Contemporary China*, Polity Press.

Guo, G. (2007): "Organizational Involvement and Political Participation in China," *Comparative Political Studies*, 40, 457–482.

Gyourko, J. and R. Molloy (2014): "Regulation and housing supply," *NBER working paper*.

Hallward-Driemeier, M. and L. Pritchett (2015): "How business is done in the developing world: deals versus rules," *Journal of Economic Perspectives*, 29, 121–140.

Harari, M. (2017): "Cities in bad shape: Urban geometry in India," *mimeo*.

Hayek, F. A. (1945): "The use of knowledge in society," *American Economic Review*, 35, 519–530.

Helpman, E. (1998): *Topics in public economics: theoretical and applied analysis*, Cambridge, UK: Cambridge Univ. Press, chap. The size of regions, 33–54.

Heurlin, C. (2016): *Responsive Authoritarianism in China: Land, Protests, and Policy Making*, Cambridge University Press.

Hibbs, Douglas A., J., R. D. Rivers, and N. Vasilatos (1982): "On the Demand for Economic Outcomes: Macroeconomic Performance and Mass Political Support in the United States, Great Britain and Germany," *Journal of Politics*, 44, 426–462.

Hopkins, D. and G. King (2010): "A Method of Automated Nonparametric Content Analysis for Social Science," *American Journal of Political Science*, 54, 229–247.

Huang, Y. (1996): "Central-local relations in China during the reform era: The economic and institutional dimensions," *World Development*, 24, 655–672.

Iyer, L., X. Meng, N. Qian, and X. Zhao (2013): "Economic transition and private-sector labor demand: Evidence from urban China," *NBER working paper*.

Jacobs, J. (1969): "The economy of cities," *New York: Vintage*.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): "Geographic localization of knowledge spillovers as evidenced by patent citations," *the Quarterly journal of Economics*, 108, 577–598.

Jiang, J., T. Meng, and Q. Zhang (Forthcoming): "From internet to social safety net: The policy consequences of online participation in China," *Governance*.

Jin, H., Y. Qian, and B. R. Weingast (2005): "Regional decentralization and fiscal incentives: Federalism, Chinese style," *Journal of Public Economics*, 89, 1719–1742.

Juhász, R. and C. Steinwender (2018): "Spinning the web: codifiability, information frictions and trade," *NBER working paper*.

Kim, S. and J. Lee (2012): "E-Participation, Transparency, and Trust in Local Government," *Public Administration Review*, 72, 819–828.

King, G., J. Pan, and M. E. Roberts (2013): "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, 107, 326–343.

——— (2017): "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument," *American Political Science Review*, 111, 484–501.

La Porta, R., F. Lopez-de Silanes, and A. Shleifer (2008): "The economic consequences of legal origins," *Journal of Economic Literature*, 46, 285–332.

Langford, E., N. Schwertman, and M. Owens (2001): "Is the Property of Being Positively Correlated Transitive?" *The American Statistician*, 55, 322–325.

Lardy, N. (2014): *Markets Over Mao: the rise of private business in China*, Peterson Institute for International Economics.

Lelkes, Y., G. Sood, and S. Iyengar (2015): "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect," *American Journal of Political Science*, 61, 5–20.

Li, H. and L.-A. Zhou (2005): "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China," *Journal of Public Economics*, 89, 1743–62.

Linde, J. and M. Karlsson (2013): "The Dictator's New Clothes: The Relationship Between E-Participation and Quality of Government in Non-Democratic Regimes," *International Journal of Public Administration*, 36, 269–281.

Lindner, R. and U. Riehm (2011): "Broadening Participation Through E-Petitions? An Empirical Study of Petitions to the German Parliament," *Policy & Internet*, 3, 63–85.

Liu, C., J. Li, and J. Liu (2015): "Rural E-commerce and New Model of Rural Development in China: A Comparative Study of "Taobao Village" in Jiangsu Province." *Asian Agricultural Research*, 7, 35 – 46.

Liu, X. and Y. Long (2016): "Automated identification and characterization of parcels with OpenStreetMap and points of interest," *Environment and Planning B: Planning and Design*, 43, 341–360.

Long, Y. (2016): "Redefining Chinese city system with emerging new data," *Applied Geography*, 75, 36–48.

Lorentzen, P. (2013): "Regularizing Rioting: Permitting Public Protest in an Authoritarian Regime," *Quarterly Journal of Political Science*, 8, 127–158.

Malesky, E. and J. London (2014): "The Political Economy of Development in China and Vietnam," *Annual Review of Political Science*, 17, 395–419.

Marien, S., M. Hooghe, and E. Quintelier (2010): "Inequalities in Non-institutionalised Forms of Political Participation: A Multi-level Analysis of 25 countries," *Political Studies*, 58, 187–213.

Marshall, A. (1890): *Principles of economics*, London: Macmillan.

McCallum, A. K. (2002): "MALLET: A Machine Learning for Language Toolkit," Http://mallet.cs.umass.edu.

Meltzer, A. and S. F. Richard (1981): "A Rational Theory of the Size of Government," *Journal of Political Economy*, 89, 914–927.

Meng, T., J. Pan, and P. Yang (2017): "Conditional Receptivity to Citizen Participation," 50, 399–433.

Mihalcea, R. and P. Tarau (2004): "Textrank: bringing order into text," in *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Mikolov, T., K. Chen, G. Corrado, and J. rey Dean (2013a): "Efficient estimation of word representations in vector space." *arxiv. org/abs/1301.3781*.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013b): "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 3111–3119.

Miller, M. K. (2015): "Elections, Information, and Policy Responsiveness in Autocratic Regimes," *Comparative Political Studies*, 48, 691–727.

Miller, W. E. and D. E. Stokes (1963): "Constituency Influence in Congress," *The American Political Science Review*, 57, 45–56.

Mimno, D., H. M. Wallach, E. Talley, M. Leenders, and A. McCallum (2011): "Optimizing Semantic Coherence in Topic Models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, EMNLP '11, 262–272.

Moretti, E. (2004): "Human capital externalities in cities," in *Handbook of regional and urban economics*, Elsevier.

Naughton, B. J. (2018): *The Chinese economy*, MIT Press Ltd.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

Norris, D. F. and C. G. Reddick (2012): "Local E-Government in the United States: Transformation or Incremental Change?" *Public Administration Review*, 73, 165–175.

Norris, P. (2001): *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*, Cambridge University Press.

Oreglia, E. (2015): "The "Sent-Down" Internet: using information and communication technologies in rural China," *Chinese Journal of Communication*, 8, 1–6.

Pan, J. (2017): "How Chinese Officials Use the Internet to Construct Their Public Image," *Political Science Research and Methods*, 1–17.

Persson, P. and E. Zhuravskaya (2016): "The Limits of Career Concerns in Federalism: Evidence from China," *Journal of the European Economic Association*, 14, 338–374.

qinwf (2016): "jiebaR: Chinese text segmentation with R," Https://github.com/qinwf/jiebaR.

Rauch, J. (1993): "Productivity gains from geographic concentration of human capital: Evidence from the cities," *Journal of Urban Economics*, 34, 380–400.

Rosenthal, S. S. and W. C. Strange (2004): "Evidence on the nature and sources of agglomeration economies," in *Handbook of regional and urban economics*, Elsevier, vol. 4, 2119–2171.

——— (2008): "The attenuation of human capital spillovers," *Journal of Urban Economics*.

Royo, S., A. Yetano, and B. Acerete (2013): "E-Participation and Environmental Protection: Are Local Governments Really Committed?" *Public Administration Review*, 74, 87–98.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011): "The area and population of cities: New insights from a different perspective on cities," *American Economic Review*, 101, 2205–25.

Saiz, A. (2010): "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 125, 1253–1296.

Schlozman, K. L., S. Verba, and H. E. Brady (2010): "Weapon of the Strong? Participatory Inequality and the Internet," *Perspectives on Politics*, 8, 487–509.

Shi, T. (1997): *Political Participation in Beijing*, Cambridge MA: Harvard University Press.

Solinger, D. J. and Y. Hu (2012): "Welfare, Wealth and Poverty in Urban China: The Dibao and Its Differential Disbursement," *The China Quarterly*, 211, 741–764.

Stromseth, J. R., E. J. Malesky, and D. D. Gueorguiev (2017): *China's Governance Puzzle: Enabling Transparency and Participation in a Single-Party State*, Cambridge University Press.

Tang, W. (2016): *Populist Authoritarianism: Chinese Political Culture and Regime Sustainability*, Oxford University Press.

Tian, Y. (2019): "International trade liberalization and domestic institutional reform: effects of WTO accession on Chinese internal migration policy," .

Torres, L., V. Pina, and B. Acerete (2006): "E-Governance Developments in European Union Cities: Reshaping Government's Relationship with Citizens," *Governance*, 19, 277–302.

Truex, R. (2016): *Making Autocracy Work*, Cambridge University Press.

Tsai, L. L. and Y. Xu (Forthcoming): "Outspoken Insiders: Political Connections and Citizen Participation in Authoritarian China," *Political Behavior*.

Turner, M. A., A. Haughwout, and W. Van Der Klaauw (2014): "Land use regulation and welfare," *Econometrica*, 82, 1341–1403.

United Nations (2016): "E-Government Survey 2016: E-Government in Support of Sustainable Development," .

Verba, S. and N. H. Nie (1972): *Participation in America: Political Democracy and Social Equality*, Evanston, London.

Vogel, K. B., R. Goldblatt, G. H. Hanson, and A. K. Khandelwal (2018): "Detecting urban markets with satellite imagery: An application to India," *NBER working paper*.

Wang, S.-Y. (2011): "State misallocation and housing prices: Theory and evidence from China," *American Economic Review*, 101, 2081–2107.

Ward, S., R. Gibson, and W. Lusoli (2003): "Online Participation and Mobilisation in Britain: Hype, Hope and Reality," *Parliamentary Affairs*, 56, 652–668.

Williamson, J. (2004): "The Washington Consensus as policy prescription for development," *Challenges in the 1990s*, 33.

Wong, W. and E. Welch (2004): "Does E-Government Promote Accountability? A Comparative Analysis of Website Openness and Government Accountability," *Governance*, 17, 275–297.

Wright, J., ed. (1971): *Industrialized building in the Soviet Union*, National Bureau of Standards Special Publication 334.

Xu, C. (2011): "The Fundamental Institutions of China's Reform and Development," *Journal of Economic Literature*, 49, 1076–1151.

Xu, M., C. He, Z. Liu, and Y. Dou (2016): "How did urban land expand in China between 1992 and 2015? A multi-scale landscape analysis," *PloS one*, 11, e0154839.

Yang, G. (2009): *The Power of the Internet in China: Citizen Activism Online*, Columbia University Press.

Zhang, H. (2017): "Sunlight distance restriction, population density and the growth of Chinese cities (in Chinese)," *China Economic Quarterly*, 17, 333–354.

Zheng, Z., Y. Su, and Y. Zhang (2018): "The social origins of public support in China and its changes: evidence from the 2015 China social governance survey (in Chinese)," *Political Science Research*, 03, 51–61.

Zhou, Y. and Y. Long (2016): "SinoGrids: A practice for open urban data in China," *Cartography and Geographic Information Science*, 43, 379–392.

Appendix A

Appendix to Chapter 1

A.1 Details of Dataset Construction

A.1.1 Land Transactions

A substantial share of the residential land transaction records mark legal procedures at local land bureaus for converting land to "transferable", as can be seen from the description of transactions. In the planning era, housing was distributed to workers as a benefit, and the housing market did not exist. In the early days of urban housing reforms when the state entitled urban households with property rights over their dwellings to establish a housing market, households were made to pay for obtaining these property rights (Iyer et al., 2013). It was not easy to persuade everyone to pay up, and those who did not pay a sufficiently high price did not enjoy full property rights - they could not sell, rent out or use as collateral their living quarters, unless they pay later to convert their land into a transferable status. To the extent that we are interested in the expansion of urban areas and construction of new residential units, we want to exclude these transactions from the sample. There is no attribute that definitively marks such transactions. As a result, I classify a transaction as a procedural conversion to transferable land if the land is brown-field and the recorded means of transfer is "by contract". This is reasonable because new residential developments are legally required to acquire land through auctions. As a validation of this classification, transactions identified as "conversion to transferable land" have a size of 0.096 hectares (10,333 square feet) at its 90th percentile, much smaller than the size of lots in the remaining sample, whose 25th percentile is 0.11 hectares (11,840

square feet). I exclude these "conversion to transferable" transactions from analysis.

A.1.2 Identifying Urban Clusters

Administrative boundaries often provide poor characterizations of the boundaries of urban economies. Rozenfeld et al. (2011) propose a City Clustering Algorithm (CCA) that defines cities as maximally connected clusters of populated sites. Dingel et al. (2018) construct cities as contiguous grid cells with nighttime lights above a threshold. Vogel et al. (2018) instead use built-up landcover inferred from daytime satellite imagery to delineate cites.

I define cities as the built environment. I start with all 1 km^2 grid cells that contain at least one road intersection. There are 456,749 such cells, accounting for around 4.8% of China's land territory. I then use DBSCAN, a widely used clustering algorithm (Ester et al., 1996), to group these cells into urban clusters. The algorithm works similarly to the CCA by maximally growing clusters through combining nearby cells within a distance threshold. In a similar effort, Long (2016) defines urban areas by clustering road intersections using a different algorithm.

Figure A1 shows urban clusters discovered within Ordos, Inner Mongolia and Shanghai respectively. In the case of Ordos, we see an isolated urban area (depicted in pink) to the southwest of Ordos' main urban area (depicted in green). Kangbashi, this outlying urban area, has received international media exposure for its vast unoccupied buildings and been dubbed a "ghost town". In the case of Shanghai, we see two separate urban areas located on the outlying islands.

I use the simple test proposed by Gabaix and Ibragimov (2011) to evaluate the conformity of cluster size distribution to Zipf's Law. Figure A3 shows that these urban clusters conform almost perfectly to Zipf's Law in both land area and population. Panel A shows land area distribution across these urban clusters, where clusters smaller than 5 km^2 are excluded. Estimated coefficient of log rank (minus 0.5) on log area is -1.07, with an R^2

of 0.99. Including clusters smaller than 5 km^2 changes goodness of fit with Zipf's Law only marginally. Panel B presents population distribution across urban clusters, where population within a cluster is aggregated up from 2010 1-km^2-cell level population data from Worldpop. Clusters containing fewer than 100,000 people are excluded following Chauvin et al. (2017). Estimated coefficient of log rank (minus 0.5) on log population is -1.04, with an R^2 of 0.98. Including the less populated clusters significantly weakens goodness of fit with Zipf's Law, as is expected. Taken together, this exercise suggests that the urban clusters I identify are meaningful units of analysis.

A.2 Value of Lifting Regulatory FAR

We saw in the text that the sunlight policy describes the cross-latitude variation in residential projects' regulatory FAR quite well. For the policy to impact urban density, these regulatory upper limits have to be actually binding. If developers never want to build as much floor space as these upper limits let them, the policy has no bite.

This can be studied by looking at how land prices respond to FAR upper limit. Brueckner et al. (2017) employs this idea to around 50,000 land transactions from China, and reports an average elasticity of price to FAR upper limit of 0.75 for residential land. Their sample is a fraction of my sample, has a different time horizon (2002-2011), and tends to concentrate in the larger cities.

I estimate the following equation:

$$\log \text{Price}_{pct} = \beta \log \text{FAR}_p + \lambda_c + \gamma_1' X_p + \gamma_2' X_t + \epsilon_{pct}$$

where the dependent variable is log of real price per square meter. λ_c is a city fixed effect. X_p and X_t are the same project-level and time-level controls as in the text. We only include lots transferred through auctions, as auctions better reflect the market value of land. This excludes the vast majority of public housing projects.

Results are presented in Table A2. Our preferred specification is in column 3 with the full set of controls, where the elasticity of price with respect to FAR upper limit is estimated at 0.60. If the upper limit is not binding at all, we would expect the elasticity to be 0. 0.60 implies a substantial degree of constraint on average. Columns 4 through 7 look at heterogeneity in the restrictiveness of regulatory FAR. We see that housing projects far from the city center are not significantly less constrained by regulatory FAR. If anything, they appear to be more constrained. This is again consistent with the perception that there is often strong demand at the edges of Chinese cities. There is not much heterogeneity with respect to city population, and cities with higher GDP per capita appear to be slightly less constrained. In column 7, we see that cities with higher population densities are also not significantly more constrained by air right regulations.

A.3 Relationship between Regulatory FAR and Latitude, Raw Data

Figure A5 plots the raw data of regulatory FAR upper limit of individual residential lots against their latitude, and fits generalized additive models to estimate conditional means of FAR given latitude without controlling for any covariates. Panel A does this for all residential land, and there appears to be a negative relationship between latitude and FAR - lots in the south accommodate more floor space. The curve dips on the left end reaching into Hainan Island (the gap above 20 degrees shows the strait between the mainland and the island). To increase confidence that this negative relationship is not induced by specific regions, panels B-D separate China into three slices by longitude, and repeat the exercise. The pattern is present in all cases.

A.4 Density in Beijing, Shanghai and Shenzhen

Urbanization rate in China grew from 17.9% in 1978 to 57.9% in 2017. The area of urban land grew by 498% between 1992 and 2015 (Xu et al., 2016). Despite this rapid expansion of urban areas across China, the urban fabric varies greatly across cities. Figure A7 shows the distribution of population, points of interest (businesses and public facilities) and road intersection density in three of China' s four largest cities: Beijing, Shanghai and Shenzhen. Though comparable in overall population and GDP per capita,[1] Shanghai is much denser than Beijing, and Shenzhen much denser than Shanghai. The median 1-km urban grid cell in Shenzhen has 0.64 log points (64%) more population than the median urban cell in Shanghai, which in turn has 1.14 log points (114%) more population than the median urban cell in Beijing. This difference in density is even greater in points of interest. The median 1-km urban cell in Shenzhen has 1.23 log points more POIs than the median urban cell in Shanghai, which in turn has 0.86 log points more POIs than its counterpart in Beijing. Turning to the density of road junctions, which characterizes block size and constitutes a most visible aspect of the streetscape, the median urban cell in Shenzhen contains 0.69 log points more road junctions than that in Shanghai. The difference between Shanghai and Beijing is smaller but still substantial, with the median urban cell in Shanghai containing 0.25 log points more road junctions than the median urban cell in Beijing. As these density plots could be sensitive to definitions of what constitute a city, I further restrict the sample to city centers, defined as 1-km grid cells with a 2013 nighttime light luminosity above 50 on a scale of 0 to 63, and plot the same kernel densities for the three cities in Figure A9. We see that the patterns largely remain, though the magnitude of differences becomes somewhat smaller, and the ranking between Shanghai and Beijing in terms of road junction density is reversed.

[1] In 2010, a census year, population of Beijing is 18.8 million, Shanghai 22.3 million, Shenzhen 10.4 million. In 2016, GDP per capita of Beijing is 118 thousand yuan, Shanghai 117 thousand yuan, Shenzhen 167 thousand yuan. These data all correspond to districts directly governed by these cities, excluding counties within their jurisdiction boundaries.

Lying on the south coast of China and bordering Hong Kong, Shenzhen is 8.3 degrees south of Shanghai in latitude, which in turn is 9 degrees south of Beijing. Although the differences in density across these three cities are certainly not entirely due to the sunlight policy, these patterns are indeed consistent with the policy playing a role.

A.5 Petition Topics

These topics are obtained from a 30-topic LDA model trained on the full set of postings on LLMB up until mid-2016 (for details of the methodology, see Jiang et al. (Forthcoming)). The topics generally have very clear meanings. I group them into urban topics and rural topics, and focus on urban issues in this paper. I further classify the urban topics into grievance topics and non-grievance topics, which is quite evident from their contents. Below I list words with highest probability weights in each topic.

A.5.1 Urban Topics, Grievances

Topic 1: residential compound, property owner, real estate property, have not, residents, problem, elevator, occupants, property management company, inside, management, property management fee, garden, agency, community

Topic 2: vehicle, road, traffic, severe, safety, passenger, crossroads, streetlight, do not have, agency, segment of road, travel, cause, influence, hope

Topic 4: garbage, pollution, severe, residents, environment, life, waste water, influence, one, nearby, hope, leader, health, emission, production

Topic 5: bus, public transport, driver, taxi, vehicle, passenger, convenience, time, have not, travel, hours, hope, traffic, car, train station

Topic 6: police station, one, pyramid selling scam, police, personnel, happen, police department, at that time, call the police, tour guide, touring, have not, hope, friends, police

Topic 7: hospitals, father, reimbursement, doctors, life, mother, treatment, family, medical insurance, children, cannot, handicapped, elderly, in-patient, expenses

Topic 8: demolishing, home, settlement, upgrade, government, house, have not, compensation, leader, my home, construction, planning, secretary, shanty-town, relocated households

Topic 11: household registration, children, handle, have not, policy, certificate, work, cannot, police branch, one, parents, leader, need, please, residence

Topic 12: house, residents, severe, safety, construction, department, remove, in the process of construction, occupants, house, problem, influence, residential compounds, have not, cause

Topic 14: operate, market, department, one, influence, severe, urban management officer, hope, manage, environment, Internet cafe, leader, secretary, gambling, commercial tenant

Topic 15: school, students, children, teacher, primary school, parents, kindergarten, education, middle school, make-up class, attend school, education bureau, leader, learn, one

Topic 17: developer, property owner, real estate transaction, apartment, have not, now, residential compound, property ownership certificate, handle, contract, government, develop, leader, already, purchase

Topic 18: test, work, college students, teacher, driving school, civil servants, have not, not yet, graduation, participate, professional, village officer, employment, one, graduating student, test taker

Topic 21: charge fees, charge, fee, regulation, standard, price, state, whether, reasonable, unreasonable charges, please, document, natural gas, request

Topic 22: company, salary, employee, firm, have not, leader, migrant worker, limited-liability company, worker, secretary, unit, arrears, projects, labor, ten thousand yuan

Topic 24: telephone, have not, handle, staff, information, Internet, company, cannot,

one, broadband, make a call, complain, bank, mobile, phone

Topic 26: salary, work, teacher, have not, unit, personnel, retirement, employee, compensation package, life, secretary, policy, leader, state, now

Topic 28: residents, noise, affect, severe, residential compounds, at night, life, disruptive, department, everyday, in the process of construction, normal, environment, one

Topic 30: residential compounds, heating, company, solve, residents, have not, heat, water supply, heat supply, occupants, leader, life, now, water supply shutdown

A.5.2 Urban Topics, Non-grievances

Topic 3: secretary, problem, solve, leader, hope, hello, not yet, respect, reflect, mayor, thank you, take time from a busy schedule, attention, now, ask

Topic 9: have not, now, one, know, leader, folks, hope, really, secretary, government, once, why not, cannot, location, see

Topic 10: planning, construction, railway, residents, convenience, wide road, traffic, highway, have not, public transport, residential compound, connect, please, nearby

Topic 19: regulation, relevant, department, undertake, require, law, illegal, condition, government, court, report, have not, state, behavior, unit

Topic 20: leader, local, People's Daily website, message board, source, secretary, hello, once, respect, hope, have not, hi, now, thank you, pay attention to

Topic 25: government, problem, mass, work, people, society, leader, folks, cadre, department, one, real, hope, proceed, should

Topic 27: develop, construction, city, economy, one, suggestion, travel, culture, hope, rural areas, people, whole country, hometown, environment, secretary

A.5.3 Rural Topics

Topic 13: villager, farmer, reservoir, severe, cultivation, cause, damage, production, government, extraction, river course, now, land, secretary, farmland

Topic 16: villager, road, path, leader, secretary, now, have not, build a road, travel, cement road, suddenly, road surface, bumpy, raining day

Topic 23: village, farmer, have not, secretary, leader, policy, state, compensation, subsidy, low-income allowance, this year, loan, one, finance, poverty alleviation

Topic 29: villager, land, have not, secretary, farmland, compensation, land taking, village committee, village cadres, occupy, my home, rural residential land, forcefully, compensation package

A.6 Longitude as a Placebo

Since density does not systematically vary over longitudes, we should expect the response speed of petitions to government reply rate not to vary across longitudes. Regressions reported in Table A3 repeat those in Table 1.8, with predicted log FAR being replaced by longitude. We see that almost all of the coefficients are insignificant, suggesting that the speed of information transmission does not vary across longitudes.

Data	Description and Citations	Time coverage	Source
Land transactions	Over 1.34 million land transaction records containing information on geographical location, price, zoned use, regulatory floor area ratio, size, etc.	2000-2016, covers 69.8% in area of all urban land transferred through the government post 2006	Ministry of Land and Resources
Road junctions	Number of road junctions within 1-km grid cells for all China, Zhou and Long (2016), Liu and Long (2016)	2011	Beijing City Lab
Population density	Population within 1-km grid cells for all China, Gaughan et al. (2016)	1990, 2000, 2005, 2010, 2015	WorldPop and Gridded Population of the World
City-level economic characteristics	GDP per capita, population, population density, share of population attending primary school, middle school, etc.	1994-2014 for prefectures, 1997-2013 for counties	China Data Center, University of Michigan
Nighttime light	Satellite-detected nighttime light at 1-km-grid-cell level for all China	1992-2013	DMSP-OLS
Ruggedness	Share of land with different slopes within 1-km grid cells, Fischer et al. (2008)		FAO Harmonized World Soil Database v1.2
Climate	Average January temperature, average July temperature, and average precipitation within 1-km grid cells, Fick and Hijmans (2017)		WorldClim v2
Population mini-census	A 1% population survey carried out by National Bureau of Statistics. Contains rich household and individual level information	2005	
Online petitions	Around 900,000 postings on a national petition platform that hosts sub-boards for all administrative units	2008-2016	Local Leader Message Board, People's Daily

Table A1: Data Sources

Table A2: Value of Loosening Regulatory FAR Upper Limit

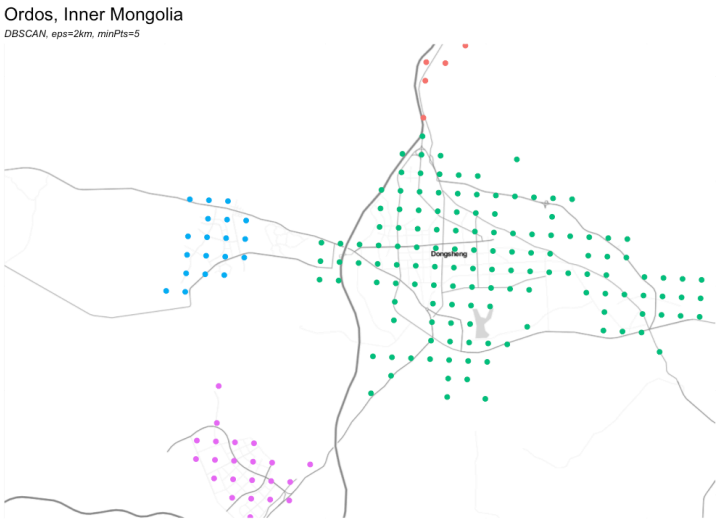
	Dependent Variable: Log Real Price per Square Meter						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log FAR upper limit	0.843***	0.743***	0.601***	0.573***	0.602***	0.638***	0.624***
	(0.042)	(0.020)	(0.017)	(0.022)	(0.033)	(0.032)	(0.033)
Distance to city center, above median			-0.269***	-0.319***	-0.265***	-0.263***	-0.269***
			(0.017)	(0.031)	(0.018)	(0.018)	(0.017)
Log FAR upper limit × Distance to city center, above median				0.057*			
				(0.029)			
Log FAR upper limit × Population, above median					-0.008		
					(0.039)		
Log FAR upper limit × GDP pc, above median						-0.071*	
						(0.038)	
Log FAR upper limit × Population density, above median							-0.036
							(0.039)
City FEs	No	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	No	Yes	Yes	Yes	Yes	Yes	Yes
Lot-specific controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	164346	164346	144241	144241	138502	138257	144241
R-Squared	0.128	0.503	0.558	0.558	0.561	0.561	0.558
RMSE	0.979	0.744	0.704	0.704	0.706	0.706	0.704

Notes: These regressions concern residential development projects across China for which the land parcel is auctioned off between 2007 and 2016. Dependent variable is log real price per square meter. Time FEs include year and month-of-year fixed effects. Lot-specific controls include whether the lot is greenfield or brown-field, type of residential building, fixed effects for land grade, whether parcel size is above median, whether distance to 2005 city center is above median and 2005 nighttime lights. City-level population and GDP per capita are 2006 levels obtained from statistical yearbooks. City-level population density is 2010 level calculated from Worldpop data. Standard errors are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

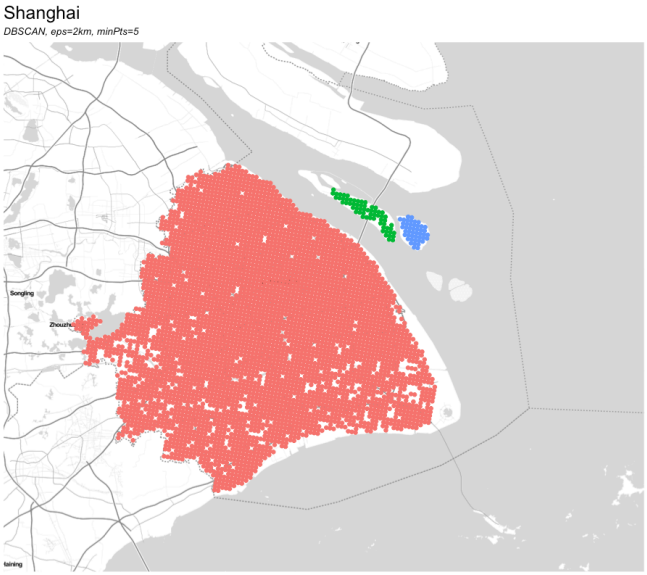
Table A3: Petition Reaction to Government Replies, Longitude Placebo

	Dependent Variable: Log Number of Posts				
	(1)	(2)	(3)	(4)	(5)
Reply rate × Longitude	-0.002	-0.002	-0.000	-0.001	-0.039**
	(0.003)	(0.003)	(0.003)	(0.004)	(0.015)
Reply rate$_{t-1}$ × Longitude	0.001	-0.002	-0.002	-0.005	-0.016
	(0.002)	(0.003)	(0.003)	(0.003)	(0.013)
Reply rate$_{t-2}$ × Longitude	0.001	-0.002	-0.001	-0.003	0.020
	(0.002)	(0.003)	(0.003)	(0.003)	(0.013)
Reply rate$_{t-3}$ × Longitude	0.004*	0.001	0.002	0.002	-0.023*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.013)
Reply rate$_{t-4}$ × Longitude	0.005*	0.003	0.003	0.003	-0.020
	(0.003)	(0.003)	(0.003)	(0.003)	(0.014)
Reply rate$_{t-5}$ × Longitude	0.001	0.000	0.002	0.001	-0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.012)
Reply rate$_{t-6}$ × Longitude	0.000	-0.004	-0.003	-0.006**	-0.020
	(0.002)	(0.003)	(0.003)	(0.003)	(0.013)
City FEs	Yes	Yes	Yes	Yes	Yes
Topic-Quarter FEs	Yes	Yes	Yes	Yes	Yes
Education interactions	No	Yes	Yes	Yes	Yes
Economic interactions	No	No	Yes	Yes	Yes
Communication interactions	No	No	No	Yes	Yes
Province FE interactions	No	No	No	No	Yes
Observations	19317	19252	18538	18176	18176
R-Squared	0.622	0.625	0.613	0.618	0.626
RMSE	0.565	0.564	0.566	0.566	0.562

Notes: Units of observation in these regressions are city-quarter-topic tuples. Topics are obtained from a 30-topic LDA model. Each post is assigned to its highest topic. Dependent variable is log number of posts about a given topic posted in a given city's Local Leader Message Board in a given quarter. Coefficients of reply rate and its lags are omitted. Education interactions are reply rate and its lags interacted with shares of six education levels of city residents. Economic interactions are reply rate and its lags interacted with city's GDP per capita, fiscal expenditure to GDP ratio, social security spending to GDP ratio and social assistance to GDP ratio. Communication interactions are reply rate and its lags interacted with the following city-level variables normalized by population: annual passengers transported, annual freight transported, number of post offices, landline users, mobile phone users, broadband internet users. Province FE interactions are reply rate and its lags interacted with province fixed effects. Standard errors are clustered at city-topic level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.



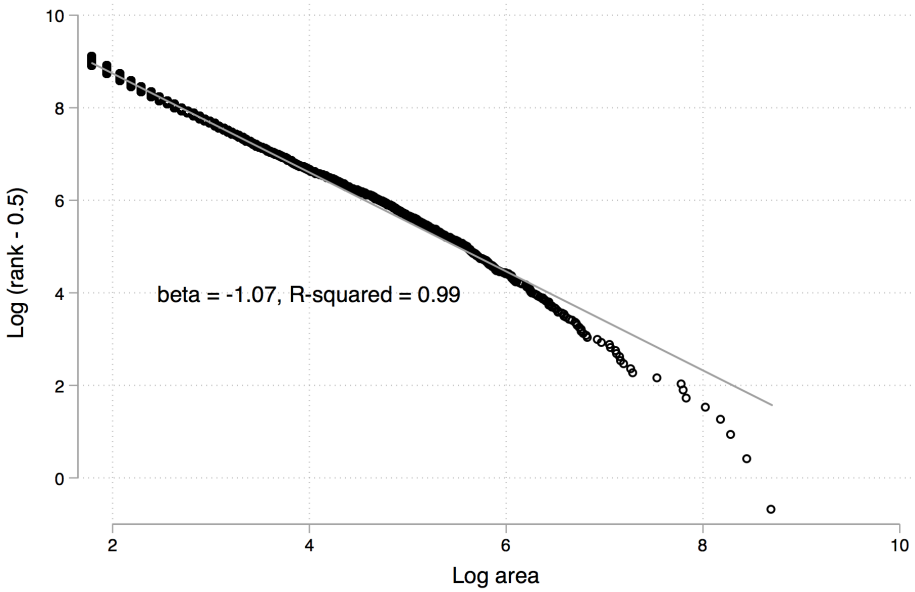
(a) Ordos



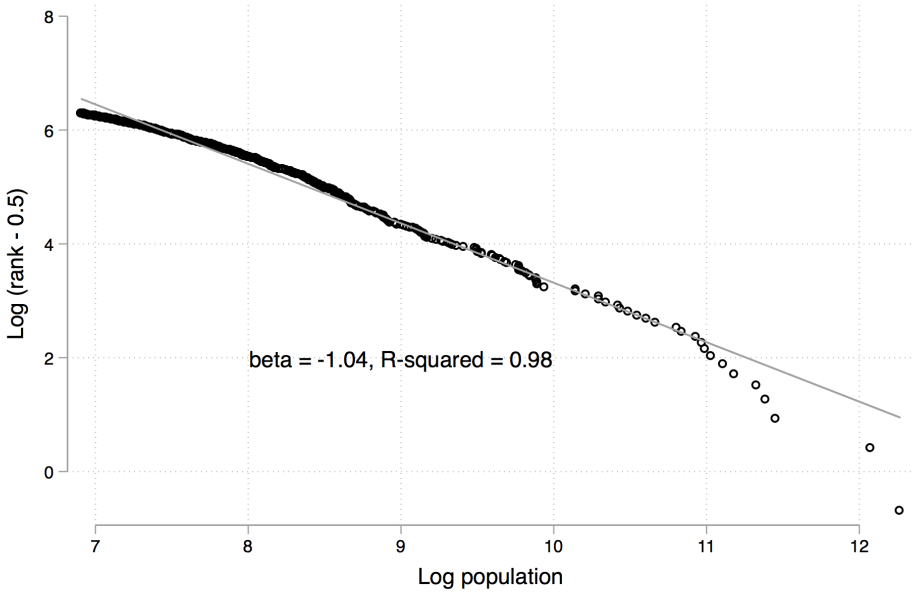
(b) Shanghai

Figure A1: Identifying Urban Clusters

Notes: These figures show urban clusters discovered with DBSCAN within jurisdicition boundaries of Ordos (Panel A) and Shanghai (Panel B). Each dot is a 1 km^2 grid cell containing at least one road intersection. Distinct colors indicate separate clusters. See details of the clustering method in text.



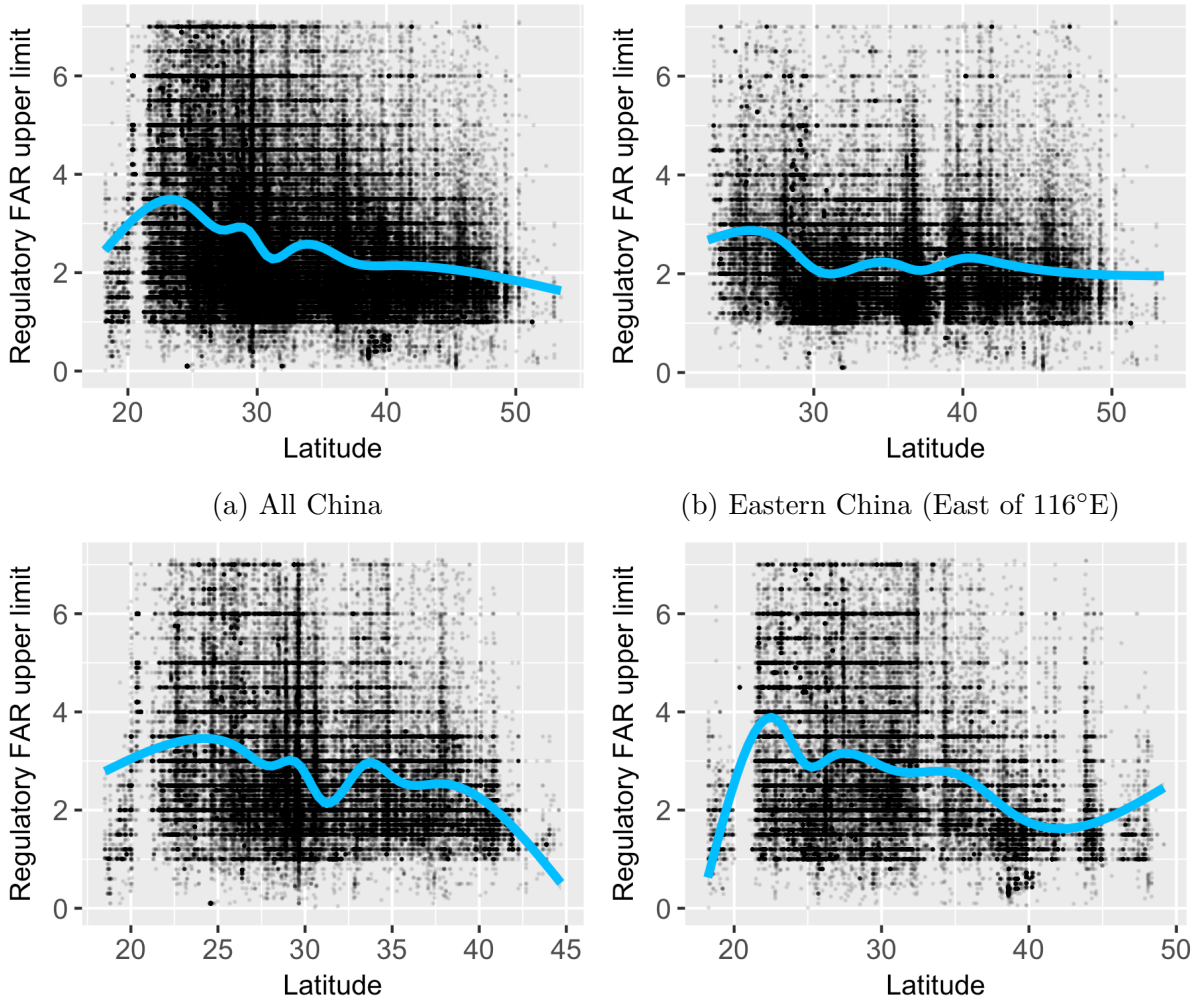
(a) Area



(b) Population

Figure A3: Zipf's Law for Urban Clusters

Notes: Panel A shows for identified urban clusters log of rank (minus 0.5) against log of cluster area, following the simple test of Zipf's Law in Gabaix and Ibragimov (2011). Urban clusters smaller than 5 km^2 are excluded from Panel A, but including them does not alter the tight Zipf fit. Panel B shows for identified urban clusters log of rank (minus 0.5) against log of cluster population in 2010, which is aggregated up from Worldpop 1 km^2 population counts. Urban clusters containing fewer than 100,000 people are excluded following the literature. Including these small towns significantly weakens the overall fit with Zipf's Law.



(a) All China

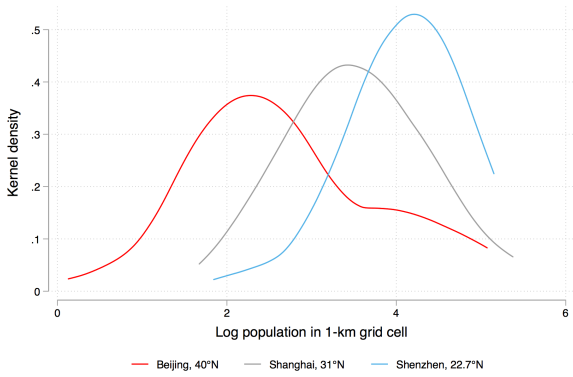
(b) Eastern China (East of 116°E)

(c) Central China (110°E-116°E)

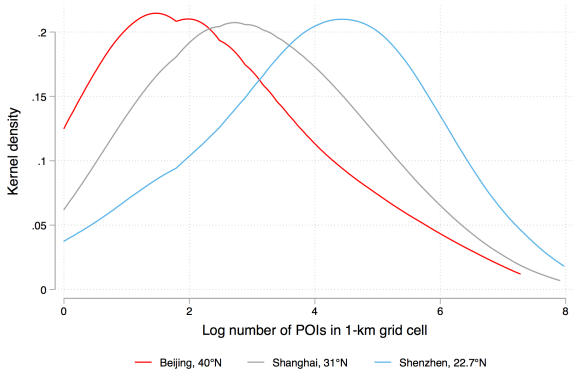
(d) Western China (West of 110°E)

Figure A5: Relationship Between Latitude and Regulatory FAR Upper Limit

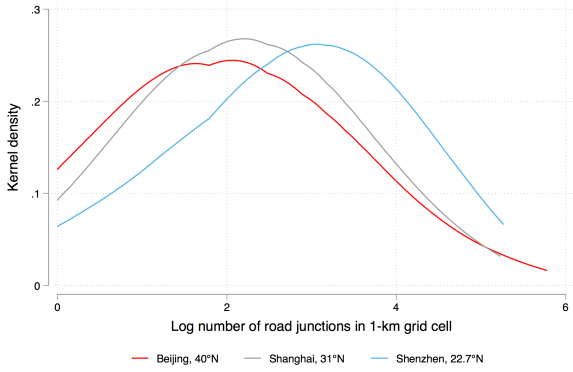
Notes: These figures are scatterplots of the latitude and regulatory FAR of individual residential lots from across China. Blue curves are generalized additive model non-parametric fits.



(a) Population



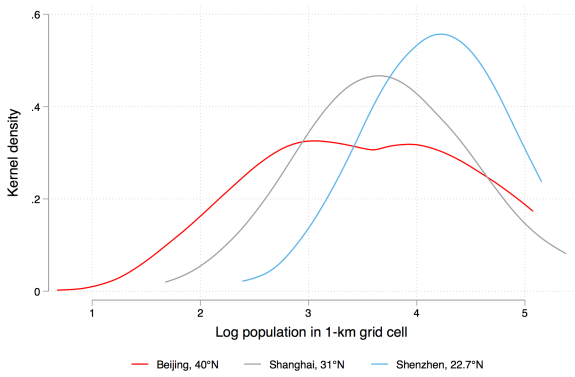
(b) Points of interest



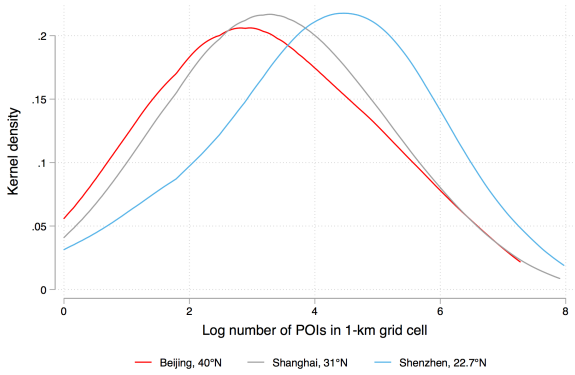
(c) Road junctions

Figure A7: Urban Density in Beijing, Shanghai and Shenzhen

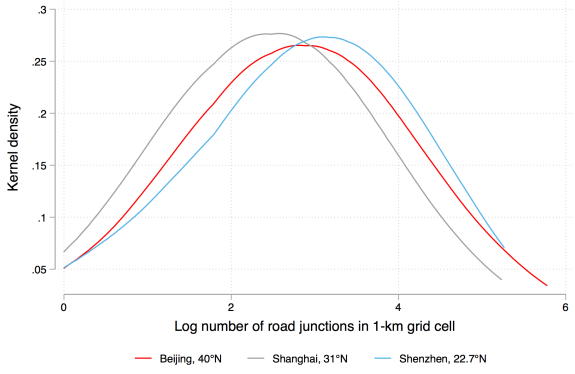
Notes: These are kernel density plots of population, the number of points of interest (businesses and public facilities) and road junctions in 1-km urban grid cells in Beijing, Shanghai and Shenzhen, where "urban" grid cells are defined as those containing at least one road junction. Only districts directly governed by these cities are included. Counties within the jurisdiction boundaries of these cities are excluded.



(a) Population



(b) Points of interest



(c) Road junctions

Figure A9: Urban Density in Beijing, Shanghai and Shenzhen, City Centers

Notes: These are kernel density plots of population, the number of points of interest (businesses and public facilities) and road junctions in 1-km grid cells that contain at least one road junction and have a 2013 nighttime light luminosity of at least 50 (on a scale of 0 to 63) in Beijing, Shanghai and Shenzhen. Only districts directly governed by these cities are included. Counties within the jurisdiction boundaries of these cities are excluded.

Appendix B

Appendix to Chapter 2

B.1 Electronic Government Petition Platforms around the World

Table B1: A List of Electronic Petition Platforms Created by National/Regional Governments

Country/region	**Creator Organization(special name, if any)**	**Year of establishment**	**Website**
United Kingdom	Scottish Parliament	1999	http://www.parliament.scot/gettinginvolved/petitions
Australia	Queensland Parliament	2002	https://www.parliament.qld.gov.au/work-of-assembly/petitions
South Korea	Korean National Government (*E-People*)	2005	http://www.epeople.go.kr/
United Kingdom	Prime Minister's Office (*Downing Street E-Petitions*)	2006-2010	http://petitions.number10.gov.uk
Germany	German Bundestag	2007	https://epetitionen.bundestag.de/epet/peteinreichen.html
China	Chinese Communist Party (*Local Leader Message Boards*)	2008	http://liuyan.people.com.cn
United Kingdom	UK Parliament	2011	https://petition.parliament.uk
United States	White House (*We the People*)	2011	https://petitions.whitehouse.gov
Russia	Russian Federal Government (*Russian Public Initiative*)	2013	https://www.roi.ru
Ukraine	Cabinet of Ministers	2016	https://petition.kmu.gov.ua

B.2 Survey Evidence on Policy Preferences across Socioeconomic Status

This section draws on data from the 2008 and 2009 China Citizenship Attitude Surveys (CCAS) to provide evidence on how Chinese citizens' policy preferences vary by their socioeconomic status. Specifically, we utilize a questionnaire item (`D10`), which asks the respondents to pick an issue that they think is the "most important" from the following 4 options:

1. Maintaining domestic order
2. Give people more voice in government decision-making
3. Protect civil liberty
4. Improve welfare of the poor

We estimate a multinomial logit model where individual choices are regressed on log family income along with other covariates. Table B2 presents the results. The results suggest that both family income and rural residency are highly significant predictors of preferred policy priorities. Compared to the urban, wealthy groups, rural and lower-income respondents are much more likely to regard welfare improvement for the poor as the most important issue. Consistent with the pattern, we also find that the probability of choosing "don't know" decreases as income rises. These results thus support our claim that the rich and the poor possess distinct preferences for policy issues in contemporary China.

Table B2: Different Demands from High- and Low-Income Individuals

	DV: Most Important Task for Government			
Baseline: public order	Voice on policy	Freedom	Improve welfare	Don't know
Log family income	0.147	-0.042	-0.140***	-0.349***
	(0.092)	(0.068)	(0.045)	(0.078)
Rural residency	-0.004	0.104	0.343***	1.190***
	(0.202)	(0.168)	(0.116)	(0.304)
Age	-0.008	0.001	0.006*	0.035***
	(0.007)	(0.005)	(0.003)	(0.006)
Female	0.143	0.168	0.455***	1.285***
	(0.152)	(0.118)	(0.075)	(0.157)
College education	-0.247	-0.170	-0.120	-14.079***
	(0.447)	(0.342)	(0.241)	(0.333)
CCP membership (self or family)	-0.103	-0.119	-0.268***	-0.793***
	(0.190)	(0.149)	(0.096)	(0.226)
Employed in state sector	-0.089	-0.108	-0.109	-0.697*
	(0.259)	(0.197)	(0.138)	(0.411)
Constant	-2.607**	-0.708	0.864*	-1.618*
	(1.073)	(0.788)	(0.511)	(0.980)
Province dummy	✓	✓	✓	✓
Year dummy	✓	✓	✓	✓
Observations	4133	4133	4133	4133

Note: This table shows results from a multinomial logit model where the dependent variable is the respondent's choice of the "most important task for government to accomplish". The result suggests that family income is negatively associated with the tendency to select welfare improvement as the most important task for the government. In other words, poor individuals care much more about social welfare issues than wealthy ones.
Data: China Citizenship Attitude Survey (2008 & 2009)
Robust standard errors clustered at individual level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.3 A Snap Shot of LLMB's Interface

B.4 Comparing Petition Activities on LLMB vs. Local Platforms

It is important to acknowledge that while the LLMB is the largest online petition platform in China during the period we analyze, it is not the only platform available. Aside from LLMB, many local governments have created their websites to receive cit-

izen complaints. This section provides some evidence on the relationship between petitions on LLMB and local platforms. We scrape petitions from three local websites where the data are publicly available: Sichuan (https://ly.scol.com.cn), Changsha (http://wlwz.changsha.gov.cn/webapp/cs/email/index.jsp) and Nanjing (http://www.njbbs.gov.cn/). Sichuan is a provincial level unit located in southwest China whereas Changsha and Hunan are two prefecture-level units, located in central and eastern parts of China, respectively. We calculate the total number of online petitions filed on these websites each day and regress it on daily petition volumes from the corresponding LLMB platforms.[1] The results are displayed in Table B3. We can see in all three localities, there exist strong and contemporaneous correlations between LLMB and local petitions. On average, one petition on LLMB is associated with about 0.4 to 1.8 petitions on local sites. These patterns provide suggestive evidence that petition activities that we observe on LLMB are broadly representative of online petition activities on local websites.

Table B3: Correlation in Petition Volumes: LLMB vs. Local Platforms

	DV: Daily Petition in Local Platforms			
	(1) All	(2) Sichuan	(3) Changsha	(4) Nanjing
Daily petitions at LLMB (t)	0.3826***	0.3836***	0.8564***	1.8410***
	(0.0223)	(0.0222)	(0.2399)	(0.5969)
Daily petitions at LLMB ($t-1$)	-0.0083	-0.0114	0.7981***	1.6483***
	(0.0243)	(0.0243)	(0.2480)	(0.6061)
Daily petitions at LLMB ($t-2$)	-0.0462**	-0.0444*	0.3729	0.1216
	(0.0235)	(0.0235)	(0.2294)	(0.6075)
Daily petitions at LLMB ($t-3$)	-0.0486**	-0.0447**	0.3297	-0.9151
	(0.0196)	(0.0195)	(0.2127)	(0.5784)
R^2	0.36	0.51	0.02	0.10
Observations	9537	3211	3365	2961

Note: This table shows the correlation in daily petition volumes between LLMB and several local petition platforms (Sichuan, Changsha, and Nanjing). The estimates show that petition activities on local platforms track closely with petitions on LLMB. Robust standard errors are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

[1]The observations are somewhat different for the three localities because the local platforms are established on different dates.

B.5 Details on Non-Parametric Content Analysis of Petition Issues and Users Backgrounds

B.5.1 Analysis Procedure

To create the training dataset, we hire two research assistants (RAs) to read through a random sample of 3,500 petitions and code each petition's issue and location. In creating the issue labels, we consulted the output from unsupervised LDA topic models (discussed in Section B.8, OA) and made additional adjustments based on a close reading of hundreds of petitions. A copy of the coding manual is available at https://www.dropbox.com/s/ufcd4imrwqbo2my/Coding%20Manual.docx?dl=0.

Once the coding of is completed, we put the training data through the HK algorithm. One practical challenge that we encountered during the implementation of the HK method, however, is that `ReadMe`, the R package that performs the estimation procedure, has a memory constraint and therefore cannot handle datasets that are too big. To circumvent this problem, we break up the data by quarter and conduct separate estimations on each smaller dataset. The proportions reported in the text are the averages of all quarterly estimates. We also inspected the quarterly distributions closely and found that they are quite stable over time.

To evaluate the quality of our estimates, we conducted ten-fold cross-validations. That is, we randomly split the training set into ten equally-sized sets, used each of the ten sets as the test set sequentially, and compared estimated proportions in the test set with true (hand-coded) proportions. The root mean squared error (RMSE) of the 10-fold cross validations is as low as 0.023 for the issue classification, meaning that an estimate of an issue proportion misses the true proportion by just 2.3% on average. The RMSE for location are slightly higher, at 0.057, but accounting for this uncertainty still does not change our substantive conclusion that a substantial share of the LLMB petitions are originated from rural and suburban areas.

B.6 Details on Estimating Government Policy Priorities

B.6.1 Methodology

We use a Latent Dirichlet Allocation (LDA) model (Blei et al., 2003) to estimate the government's attention to different policy areas in the work reports. As we will show, the algorithm discovers highly substantive and coherent topics, and measurement errors in estimated topic proportions are likely to be small. The advantages of estimating topic models over human reading and coding in our case are threefold. First, the algorithm clusters words strictly according to their co-occurrence patterns, thus avoiding the arbitrariness and errors in hand coding. The somehow ambiguous boundary between topics and the fluid nature of language makes it challenging for a human coder to consistently parse the text with some pre-specified rules, where such rules may be hard to define in the first place. An LDA can not only group together words with similar semantics, but also words conforming to similar wording styles. This later type of distinction is particularly elusive to human eyes. Second, many policies are multi-faceted, and can be attributed to more than one areas. For example, a discussion of building infrastructure is related to both the economic goal of promoting GDP growth and the welfare goal of improving people's living conditions. In this situation it is not immediately clear how one should code this piece of discussion. LDA solves this issue by allowing both topics to give rise to the word *Infrastructure*, albeit possibly with different probabilities. Which topic a particular occurrence of *Infrastructure* belongs to is then obtained via Bayesian estimation. Last but not least, using an automated algorithm tremendously reduces the cost of parsing thousands of lengthy and dry policy documents, making them a new source of data for quantitative analysis.

One might not want to take a topic model literally. That is, the data generating process of a topic model could be very different from the way these reports are actually

written. However, LDA, the most basic form of probabilistic topic models, has been shown to exhibit very good performance in a wide range of applications. Blei (2012) provides a survey of the fruitful applications of LDA in political science, psychology, population genetics, computer vision, etc. Our corpus of government work reports turns out to be a particularly good testing field for LDA, because the formal and precise wording in the reports greatly reduces noise. As a result, the topics are well-demarcated, and the majority of them are substantive (policy relevant) topics instead of "wording style" clusters.

B.6.2 Estimation Procedure

We first carry out *word segmentation* on the government work reports. Chinese characters are not naturally separated from each other as in many Indo-European languages. Therefore one has to start with segmenting and demarcating the text. We apply the *Jieba* segmentation module (qinwf, 2016) to the corpus, which uses a maximum probability segmentation model and a Hidden Markov Model (HMM) to do the segmentation. The algorithm combines an existing dictionary of Chinese words and the ability to learn new words from the text. The *Jieba* package has a proven record in word segmentation, and is 5-20 times faster than other packages. We examined the segmented text, and confirmed that the segmentation quality is very good. Special terms such as *Three Represents*, *Deng Xiaoping Theory* and *18th National Congress of the Communist Party of China* are correctly identified as a single word. The segmented text is naturally tokenized, and each piece becomes a token in subsequent computation. An unintended advantage of word segmentation is that our tokens are meaningful phrases instead of simple unigrams. Typical tokens are like *reform and open up*, *urban and rural residents* and *poverty alleviation and development*. These phrases clearly contain more information than unigrams.

We remove all punctuation, numbers and English words (such as *GDP*) from the text to focus on the Chinese vocabulary. We also remove a standard list of Chinese stop words

- common words that are not really meaningful. These include words such as *some*, *both* and *why*. Because there is no inflection in Chinese, one does not need to stem the text (i.e., reducing words to their root forms). Each word in our vocabulary denotes a unique meaning. The precise and informative nature of the government reports further adds to the information contained in the preprocessed text.

Due to properties of the Dirichlet distribution, the algorithm tends to spread a topic across few words and a document across few topics. Therefore LDA is able to find topics of much higher quality when we define each paragraph of the reports as a document $\mathbf{w}$, instead of defining each report as a document, as a paragraph is more likely to focus on a single issue. We train LDA on reports spanning from the year 2000 to the year 2013. These reports contain 440,202 paragraphs and 3,426,528 tokens in total. The LDA is thus trained on 440,202 documents. Once the model is trained with paragraphs as documents, we infer topic proportions in whole reports with Gibbs-sampling style re-sampling.

The data generating process of LDA is as follows.

A document $\mathbf{w} = (w_1, ..., w_N)$ contains N words. Each word w_i is a member of the vocabulary of V words $\{1, ..., V\}$. There are k topics, $(z_1, ..., z_k)$. k has to be specified by the modeler. Each topic is a distribution over the words. These are characterized by a $k \times V$ matrix β, where $\beta_{ij} = p(w^j = 1|z^i = 1)$.

Step 1: The term distribution β is drawn for each topic from a Dirichlet distribution with parameter δ

$$\beta \sim \text{Dirichlet}(\delta)$$

Step 2: The topic proportions of document $\mathbf{w}$ are drawn from a Dirichlet distribution with parameter α

$$\theta \sim \text{Dirichlet}(\alpha),$$

so that $\theta = (\theta_1, ..., \theta_k)$, where θ_i is the proportion of topic i.

Step 3: For each of the N words w_n,

(a) Choose a topic $z_n \sim$ multinomial(θ).

(b) Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

One needs to specify the number of topics for the model. In the following analysis, we use results from both 10 topics and 20 topics. These are reasonable numbers of policy areas generally discussed in a report. We use the Mallet program, developed by McCallum (2002), to estimate LDA. Estimation proceeds through Gibbs sampling. We monitor Markov Chain convergence in all cases. The chains generally converge before 1,000 iterations. LDA performed very well in generating meaningful clustering of words, and some of these clusters would not be easily detected by a human coder. The vast majority of the topics consist of highly coherent words, and are easily interpretable. The complete term distribution of each topic is available upon request. Section B.6.4 provides the twenty highest-probability words for each topic of the 10-topic LDA, where we have named each topic. Top words for the 20-topic LDA are also available upon request.

We use estimated topic proportions for each report $\hat{\theta}_d$ as the key independent variable for our main analysis. In the context of Gibbs sampling, these are simply the proportions of words assigned to each topic in a given report in the stationary distribution of the Markov Chains.

B.6.3 Post-Estimation Diagnostics

We examine the posterior distribution of the topic proportions to gauge the amount of noise in them. The theory of MCMC implies that the posterior distribution is the stationary distribution of the Markov Chain. Of the 1,000 sampling iterations, we discard the first 300 iterations as burn-in period, and use a thinning interval of 50. That is, starting from the 300th iteration, we take one sample every 50 iterations. In each sample, we calculate the proportion of words assigned to each topic in the whole corpus. This

yields fifteen data points for topic proportions in the whole corpus. Figure B2 plots these points. The proximity of these points for each given topic is striking, to the extent that they often look like one point. In any case, the range in proportion is smaller than 0.005. This implies that the posterior distribution is very tight, and our estimated topic proportions are very precise. To the extent that the topics are meaningful encapsulations of policy, measurement errors in policy priorities are likely to be very small.

We perform several post-estimation diagnostics to evaluate the qualities of the topics, focusing on two metrics. The first is *coherence*, which measures the tendency for top words in a topic to appear together (Mimno et al., 2011). It is defined as

$$\sum_{i} \sum_{j<i} \log \frac{N(w_j, w_i) + \beta}{N(w_i)}$$

where w_i is the ith ranked word in a given topic, $N(w_i)$ is the number of documents that contain w_i, $N(w_j, w_i)$ is the number of documents that contain both w_j and w_i, and β is a smoothing parameter. A higher coherence indicates that words in this topic are more likely to co-occur. The second one is *specificity*, which measures the distinctiveness of the words in a given topic in comparison to a uniform distribution of words, as measured by the Kullback-Leibler divergence between the word distribution of a given topic and the uniform distribution. The more specific a topic is, the greater weight it puts on a somewhat unique set of words (rather than just being a general "background" topic).

Figure B3 displays the relative positions of the 10 estimated topics in terms of both coherence (x axis) and specificity (y axis). It turns out that the welfare topic is ranked among the highest in both dimensions. This suggests that the welfare topic is one of the best quality topics generated by the 10-topic LDA model.

B.6.4 Estimated Government Work Report Topics

Topic 1: business attraction and industrial development

attract businesses, high technology, further, development zones, hundred million US dollars, value added, industrial parks, utilize foreign capital, open up, infrastructure, small and medium enterprises, emerging industries, competitiveness, develop vigorously, manufacturing, commercial zones, industrialization, major projects, focus projects, service industry

Topic 2: infrastructure and urban development

infrastructure, highway, further, overall planning, treatment plant, ten thousand square meters, project construction, environmental protection, urbanization, preliminary work, ecological environment, small cities, square kilometers, major projects, integration, comprehensive management, coverage rate, construction project, ecological improvement, polluted water treatment

Topic 3: culture, mandatory education, and public health

further, compulsory education, family planning, spiritual advancement, public health, health care, carry out thoroughly, primary and secondary schools, broadcasting system, service system, develop vigorously, culture industry, carry out extensively, institutional reform, campaign of creating, socialism, mass-line, community health, women and children

Topic 4: general economic indicators

gross product, fixed assets, net income, disposable, consumption goods, fiscal revenue, total retail sales, urban residents, value added, economic and societal, unemployment rate, people's government, growth rate, total income, provincial government, democratic parties, percentage points, household consumption, all sectors of society, rank and file of the armed police

Topic 5: real estate and high valued-added service industry

service industry, tourism, develop vigorously, further, tertiary industry, real estate, culture industry, ten thousand people, logistics industry, total income, infrastructure, value added, real estate business, e-commerce, logistics parks, home and abroad, financial

institutions, wholesale markets, consumption goods, informatization

Topic 6: political litany

economic and societal, views on development, further, Three Represents, implementation and solidification, Deng Xiaoping Theory, Eleventh Five-Year, socialism, moderately prosperous society, modernization, provincial government, Twelfth Five-Year, liberalize thoughts, reform and open up, industrialization, urbanization, structural adjustment, deepen the reforms, great standard-bearer, CPC Central Committee

Topic 7: political institutions

further, administer by law, accountability system, representatives of People's Congress, Standing Committee, democratic supervision, civil servants, carry out thoroughly, enhance clean politics, democratic parties, CPPCC members, Federation of Industry and Commerce, functions of the government, public service, people's associations, independents, party ethos and clean politics, democracy and law, public interest, establish and improve

Topic 8: agriculture modernization

agricultural produce, commercialization, leading agricultural businesses, develop vigorously, modern agriculture, standardization, further, infrastructure, herding, labor force, organic, structural adjustment, poverty alleviation and development, agriculture and herding, demonstration zone, cooperatives, production capacity, farmers and herders, farmer income, large-scale

Topic 9: market reform

further, institutional reforms, public safety, state-owned enterprises, non-public ownership, state-owned assets, public sector organizations, small and medium enterprises, crack down ruthlessly, comprehensive management, carry out thoroughly, accountability system, investment and financing, financial institutions, private firms, management system, establish and improve, letters and visits affairs, various reforms, misdemeanor and felony

Topic 10: social welfare

social safety net, endowment insurance, medical insurance, urban and rural residents, urban residents, further, labor force, ten thousand square meters, social security, coverage, migrant workers, new rural, affordable, the disabled, pension, people with financial difficulties, low income, graduates, unemployment rate, urban employees

B.6.5 A Sample of GWR with Welfare Topics Highlighted

The following paragraph is taken from the 2009 GWR of Luoyang city, with key words under the social welfare topic highlighted in red.

> [All levels and departments should]...carefully work on the issue of employment and **reemployment**, and fully implement supporting policies such as **social security** subsidies and tax breaks, with the goal of making 200 million yuan of **microfinance loans** in the whole year, adding 100 thousands urban jobs, and developing 1000 new **commonweal** posts...[We need to] encourage **migrant workers** to return home to set up businesses and to perfect the urban and rural **social security** system...[We also need to] strive to become a pilot city of rural **pension**, expedite the development of a **social security** system for **landless farmers**...[and] increase the construction of **affordable housing**.

B.7 Geographic Distribution of Online Petitions

Figure B4 illustrates the geographic variations in the average number of petitions per 10,000 residents during the sample period. Interestingly, consistent with the pattern at the individual level, the regional variations in aggregate participation also appear to be negatively associated with economic wealth: The more developed coastal areas actually saw lower per capita participation compared to many less developed, interior regions.

B.8 Details on Estimating and Classifying Online Petitions using LDA Models

B.8.1 Estimating Petition Topics

In Section B.8, we use `readMe` to infer the aggregate proportions of different issues on LLMB. However, our empirical analysis also requires classification of individual petitions. which `readMe` could not provide. To remedy this issue, we once again turn to LDA models. We fit 20-, 30-, and 40-topic LDA models to the corpus of petitions, treating each petition as a document. We discard words that contain just one Chinese character, as these are predominantly stop words. Section B.8.2 presents top words of a 30-topic LDA model. While the list of top words gives an intuitive impression that the vast majority of topics are on substantive issues (rather than specific language usage patterns such as local slang or polite formulas) and are well-separated from each other, we also present some quantitative measures supporting the high quality of the topics. We note that these topics appear to be more meaningful than those in a typical LDA application on the English language, partly due to the succinct and formal nature of the written Chinese language.

For each topic, we calculate its "exclusivity", which measures the extent to which top words for a topic are not top words for other topics. This is essentially a measure of how well topics are separated from each other. Formally, it is defined as the average (over 100 top words) of the ratio between probability of a word given a topic and sum of the probabilities of that word in all topics. For the 30-topic model, the minimum of 30 exclusivity scores is 0.1363, and the maximum is 0.4932. Even the minimum is much larger than the case of identical topics, in which case each exclusivity score would be 1/30. This indicates that the estimated topics are highly distinct from each other and can identify different types of petitions.

Since we define each petition's maximum topic as the petition's topic, we want to confirm that petitions do have dominant topics that indicate their types. Figure B5

shows the ratio between largest topic proportion and second-largest topic proportion for all petitions. It can be seen that this ratio is generally large, and has a long right tail. The mean of the distribution is 5.55, and the median is 2.00. This confirms that the largest topic in a petition does capture the main issue being discussed.

B.8.2 Details about Topic Classification

Rural Pocketbook (RP)

Topic 16: villager, road, path, leader, secretary, now, have not, build a road, travel, cement road, suddenly, road surface, bumpy, raining day

Topic 23: village, farmer, have not, secretary, leader, policy, state, compensation, subsidy, low-income allowance, this year, loan, one, finance, poverty alleviation

Topic 29: villager, land, have not, secretary, farmland, compensation, land taking, village committee, village cadres, occupy, my home, rural residential land, forcefully, compensation package

Urban Pocketbook (UP)

Topic 7: hospitals, father, reimbursement, doctors, life, mother, treatment, family, medical insurance, children, cannot, handicapped, elderly, in-patient, expenses

Topic 8: demolishing, home, settlement, upgrade, government, house, have not, compensation, leader, my home, construction, planning, secretary, shanty-town, relocated households

Topic 11: household registration, children, handle, have not, policy, certificate, work, cannot, police branch, one, parents, leader, need, please, residence

Topic 22: company, salary, employee, firm, have not, leader, migrant worker, limited-liability company, worker, secretary, unit, arrears, projects, labor, ten thousand yuan

Topic 26: salary, work, teacher, have not, unit, personnel, retirement, employee, compensation package, life, secretary, policy, leader, state, now

Rural Non-Pocketbook (RN)

Topic 13: villager, farmer, reservoir, severe, cultivation, cause, damage, production, government, extraction, river course, now, land, secretary, farmland

Urban Non-Pocketbook (UN)

Topic 1: residential compound, property owner, real estate property, have not, residents, problem, elevator, occupants, property management company, inside, management, property management fee, garden, agency, community

Topic 2: vehicle, road, traffic, severe, safety, passenger, cross-road, streetlight, do not have, agency, segment of road, travel, cause, influence, hope

Topic 3: secretary, problem, solve, leader, hope, hello, not yet, respect, reflect, mayor, thank you, take time from a busy schedule, attention, now, ask

Topic 4: garbage, pollution, severe, residents, environment, life, waste water, influence, one, nearby, hope, leader, health, emission, production

Topic 5: bus, public transport, driver, taxi, vehicle, passenger, convenience, time, have not, travel, hours, hope, traffic, car, train station

Topic 6: police station, one, pyramid selling scam, police, personnel, happen, police department, at that time, call the police, tour guide, touring, have not, hope, friends, police

Topic 9: have not, now, one, know, leader, folks, hope, really, secretary, government, once, why not, cannot, location, see

Topic 10: planning, construction, railway, residents, convenience, wide road, traffic, highway, have not, public transport, residential compound, connect, please, nearby

Topic 12: house, residents, severe, safety, construction, department, remove, in the process of construction, occupants, house, problem, influence, residential compounds, have not, cause

Topic 14: operate, market, department, one, influence, severe, urban management officer, hope, manage, environment, Internet cafe, leader, secretary, gambling, commercial tenant

Topic 15: school, students, children, teacher, primary school, parents, kindergarten, education, middle school, make-up class, attend school, education bureau, leader, learn, one

Topic 17: developer, property owner, real estate transaction, apartment, have not, now, residential compound, property ownership certificate, handle, contract, government, develop, leader, already, purchase

Topic 18: test, work, college students, teacher, driving school, civil servants, have not, not yet, graduation, participate, professional, village officer, employment, one, graduating student, test taker

Topic 19: regulation, relevant, department, undertake, require, law, illegal, condition, government, court, report, have not, state, behavior, unit

Topic 20: leader, local, people's daily website, message board, source, secretary, hello, once, respect, hope, have not, hi, now, thank you, pay attention to

Topic 21: charge fees, charge, fee, regulation, standard, price, state, whether, reasonable, unreasonable charges, please, document, natural gas, request

Topic 24: telephone, have not, handle, staff, information, Internet, company, cannot, one, broadband, make a call, complain, bank, mobile, phone

Topic 25: government, problem, mass, work, people, society, leader, folks, cadre, department, one, real, hope, proceed, should

Topic 27: develop, construction, city, economy, one, suggestion, travel, culture, hope, rural areas, people, whole country, hometown, environment, secretary

Topic 28: residents, noise, affect, severe, residential compounds, at night, life, disruptive, department, everyday, in the process of construction, normal, environment, one

Topic 30: residential compounds, heating, company, solve, residents, have not, heat, water supply, heat supply, occupants, leader, life, now, shutdown water supply

B.9 Data on Collective Protests

The Collective Incidents Dataset (CID) is a dataset compiled by the institute of sociology at the Chinese Academy of Social Sciences (CASS), one of the major state-sponsored think tanks in China. The project was motivated by the government's concern about the rising social instability since 2000. Between 2009 and 2013, the project collected on a daily basis information about collective protests and other major mass incidents,[2] and constructed a GIS based dataset, which allows researchers to analyze the spatial distribution and diffusion of mass incidents. Information about protests came from a variety of sources, including internal government documents, state-owned and commercial newspapers, TV reports, websites and social media. Extensive search was conducted on both domestic and foreign sources to minimize potential sampling bias. The dataset contains a rich set of features regarding protests recorded, such as the course of the event, location, actors, and interaction between actors, causes and consequences, process, number of participants, duration, and the information sources. While this dataset obviously does not exhaust all protests that happened in China during this period, it does provide a decent coverage on the large-scale protests, which are most likely to confound our results by inducing substantive policy changes. For the period of interest (2008-2013), the dataset contains 841 protests, with over 85% of them involving more than a thousand participants. The project was terminated in December 2013.

The other important data source is the *China Strikes* website, which records instances of labor unrest in China based on a variety of sources, including reports from website visitors. This website is dedicated to covering only contentious, collective actions by *workers over workplace issues.* Although this dataset is limited in the scope of protests covered, one of its key advantage is that it is maintained by researchers outside mainland China[3]

[2]Retrospective collection was also conducted on protests that took place prior to the project start date.

[3]Currently, this website is maintained by Manfred Elfstrom, a postdoctoral fellow at the Kennedy School of Harvard University.

and therefore less susceptible to censorship issues. For each entry, information about the date, time, location, issues, and identity of participants are provided. Sometimes, an entry will also contain photos and video clips that verify the occurrence of a protest. Between 2008 and 2012,[4] this dataset contains information for a total of 721 incidences.

We combine all the unique records from both sources into a new dataset, which contains a total of 1268 protests between 2008 and 2013. Most notably, even though these two datasets are supposedly collected by different teams and through different methods, we find a correlation of 0.74 between the two in terms of aggregate number of protests at the city level. This large correlation gives us confidence that there is a common underlying pattern of social instability that both datasets have managed to capture. In the combined dataset, a city experiences about 0.7 protests per year with a large standard deviation of 2.4. A simple correlation test suggests there is indeed a positive and significant association between the frequency of offline protest and the volume of online petitions ($\rho = 0.23$, $p < 0.0001$).

B.10 Validating the Link between Petition and Rural Participants

One potential concern with this procedure is whether our classification of LDA topics can accurately reflect participants' backgrounds. In this section, we conduct several additional tests to assess the validity of our LDA classifications. Specifically, we hypothesize that if rural petitions are indeed filed by those residing in the countryside, the volume of such petitions should be positively correlated with the size of the rural sector. To test this, we regress rural petition volumes on two measures of local rural sector size—the size of rural population and the share of rural area in the whole city. The results are shown in B4. The first two columns show that total petitions are positively correlated with both

[4]The latest data available on the website are at December 2012.

rural population size and the share of rural area. This is consistent with our claim that the LLMB is a tool that is more frequently used by rural citizens. Columns 3 through 6 further show that rural sector size is strongly and positively associated with the volumes of both total rural petitions and rural pocketbook petitions more specifically. By contrast, urban population size has no relationship with these two types of petitions. Overall, these patterns suggest that rural petitions are indeed more common in localities with a larger rural sector and support the empirical validity of our classification decisions.

Table B4: Validating the Link between Online Petitions and Rural Participants

	Total Petitions (Log)		Rural Petitions (Log)		Rural Pocketbook Petitions (Log)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log rural population (2005 population survey)	0.3383***		0.6279***		0.6388***	
	(0.0576)		(0.0578)		(0.0586)	
Log urban population (2005 population survey)	0.0845		-0.0876		-0.1053	
	(0.1052)		(0.1116)		(0.1131)	
% of rural area		2.1245*		8.9996***		8.6351***
		(1.1899)		(2.4714)		(2.4014)
Log GDP	0.3559***	0.5236***	0.1313	0.3711***	0.1450	0.3749***
	(0.0871)	(0.0656)	(0.0899)	(0.0703)	(0.0907)	(0.0708)
Year fixed-effects	✓	✓	✓	✓	✓	✓
R^2	0.37	0.26	0.42	0.25	0.41	0.23
Number of Cities	304	281	304	281	304	281
Observations	1821	1652	1821	1652	1821	1652

Note: This table shows that the volumes of both overall petitions and rural-specific petitions are positively correlated with the size of the rural sector in a locality. We use two different measures for rural sector size: rural population and rural area share. Robust standard errors clustered at city level are reported in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.11 Baseline Results with All Coefficients

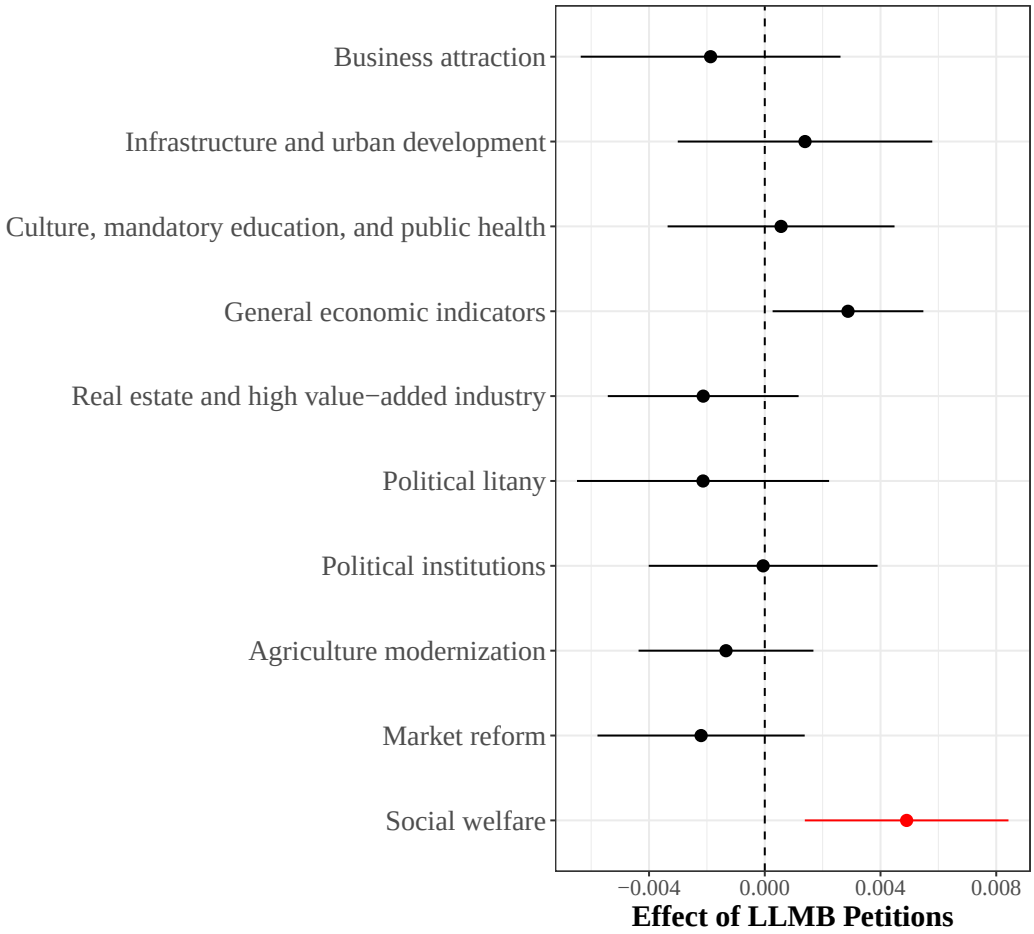
Table B5: Baseline Results with Coefficients for All Control Variables

	Δ Welfare Topic at $t+1$ (10-topic)				
	(1)	(2)	(3)	(4)	(5)
Log petitions	0.0048**				
	(0.0021)				
Log petitions (pocketbook)		0.0085***			
		(0.0029)			
Log petitions (non-pocketbook)		-0.0025			
		(0.0028)			
Log petitions (rural pocketbook)			0.0073***	0.0075***	0.0076***
			(0.0021)	(0.0022)	(0.0022)
Log petitions (urban pocketbook)			0.0019	0.0022	0.0021
			(0.0022)	(0.0023)	(0.0023)
Log petitions (rural non-pocketbook)			-0.0005	-0.0004	-0.0005
			(0.0016)	(0.0016)	(0.0016)
Log petitions (urban non-pocketbook)			-0.0023	-0.0026	-0.0026
			(0.0028)	(0.0028)	(0.0029)
Log protests				-0.0046**	-0.0047**
				(0.0023)	(0.0023)
Log employment (10,000 persons)				0.0063	0.0063
				(0.0044)	(0.0044)
Log GDP				0.0004	0.0011
				(0.0100)	(0.0102)
Log population				-0.0178**	-0.0170**
				(0.0083)	(0.0082)
GDP growth rate				-0.0006**	-0.0006**
				(0.0003)	(0.0003)
Log fiscal expenditure				0.0031	0.0025
				(0.0099)	(0.0100)
Log fiscal revenue				-0.0061	-0.0060
				(0.0065)	(0.0067)
Mayor's age					0.0000
					(0.0003)
City secretary's age					0.0002
					(0.0003)
City secretary's tenure					-0.0001
					(0.0005)
Mayor's tenure					-0.0009
					(0.0006)
City secretary's years of local service					-0.0002
					(0.0001)
Mayor's years of local service					0.0001
					(0.0002)
Connected city leader					0.0000
					(0.0022)
City and year fixed-effects	✓	✓	✓	✓	✓
Province-specific trends	✓	✓	✓	✓	✓
R^2	0.02	0.03	0.03	0.04	0.04
Number of Cities	299	299	299	298	297
Observations	1656	1656	1656	1625	1624

Note: This table shows the baseline OLS results (Table 2.2) with coefficient estimates for all the controls. Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.12 Effect of Petition on All GWR Policy Topics

Figure B6: Effect of Petitions on All GWR Topics



Note: This figure visually presents the OLS results from regressing LLMB petitions on all ten GWR topics. The horizontal lines represent the 90% confident intervals. The model is based on Column 1 of Table 2.2.

B.13 Robustness Checks

B.13.1 Baseline Estimates without City Fixed-Effects

Table B6: Baseline Estimates without Fixed-Effects

	Δ Welfare Topic at $t+1$ (10-topic)				
	(1)	(2)	(3)	(4)	(5)
Log petitions	0.0007*				
	(0.0004)				
Log petitions (pocketbook)		0.0030**			
		(0.0015)			
Log petitions (non-pocketbook)		-0.0021			
		(0.0014)			
Log petitions (rural pocketbook)			0.0021**	0.0038***	0.0039***
			(0.0010)	(0.0014)	(0.0014)
Log petitions (urban pocketbook)			0.0009	0.0008	0.0007
			(0.0013)	(0.0015)	(0.0015)
Log petitions (rural non-pocketbook)			-0.0006	-0.0004	-0.0006
			(0.0010)	(0.0011)	(0.0011)
Log petitions (urban non-pocketbook)			-0.0016	-0.0019	-0.0019
			(0.0014)	(0.0015)	(0.0015)
Year fixed-effects	✓	✓	✓	✓	✓
Economic controls				✓	✓
Leadership controls					✓
R^2	0.01	0.01	0.01	0.02	0.02
Number of Cities	299	299	299	298	297
Observations	1656	1656	1656	1625	1624

Note: This table shows regression results that do not include city or province fixed effects. The specifications are otherwise identical to those in Table 2.2.
Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.13.2 Alternative Dependent Variable: GWR Welfare Topic from a 20-Topic Model

Table B7: Using Welfare Topic from 20-Topic LDA Model as the Dependent Variable

	Δ Welfare Topic at $t+1$ (20-topic)				
	(1)	(2)	(3)	(4)	(5)
Log petitions	0.0049***				
	(0.0019)				
Log petitions (pocketbook)		0.0085***			
		(0.0025)			
Log petitions (non-pocketbook)		-0.0024			
		(0.0025)			
Log petitions (rural pocketbook)			0.0067***	0.0069***	0.0070***
			(0.0018)	(0.0019)	(0.0019)
Log petitions (urban pocketbook)			0.0025	0.0029	0.0028
			(0.0020)	(0.0021)	(0.0021)
Log petitions (rural non-pocketbook)			-0.0002	-0.0002	-0.0002
			(0.0014)	(0.0015)	(0.0015)
Log petitions (urban non-pocketbook)			-0.0024	-0.0027	-0.0026
			(0.0025)	(0.0025)	(0.0026)
City and year fixed-effects	✓	✓	✓	✓	✓
Province-specific trends	✓	✓	✓	✓	✓
Socioeconomic controls				✓	✓
Leadership controls					✓
R^2	0.03	0.03	0.04	0.04	0.04
Number of Cities	299	299	299	298	297
Observations	1656	1656	1656	1625	1624

Note: This table shows the results from OLS regressions. The dependent variables are incremental increase in the share of social welfare topic in government work reports (from 20-topic LDA model). The specification is based on Table 2.2. Robust standard errors clustered at city level are reported in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.13.3 Estimation on Different Samples

To ensure the robustness of our main findings, we re-estimate the baseline regression using several different samples. First, we add back the western provinces to see if doing so would significantly change the result. As shown in Table B8, some estimates become a bit noisier but the main patterns are substantively unchanged. Next, we split the sample by three main geographical regions: east, central, and west. The first three columns of Table B9 show that the coefficients are roughly comparable across different regions: they

are most precisely estimated in east and central regions, but also have a large, albeit noisy estimate, for the western regions. Finally, we check if our results are driven by a small number of cities with special political status. In Column 4, we exclude the provincial capitals, which are large and politically important cities. In Column 5, we exlude cities with large presence of ethnic minorities (i.e., the autonomous cities). In both cases, our main OLS results remain positive and statistically significant.

Table B8: Sub-Sample Analysis: Including Low Petition Provinces

	Δ Welfare Topic at $t+1$	
	(1)	(2)
Log petitions	0.0045**	
	(0.0020)	
Log petitions (rural pocketbook)		0.0075***
		(0.0021)
Log petitions (urban pocketbook)		0.0022
		(0.0023)
Log petitions (rural non-pocketbook)		-0.0006
		(0.0016)
Log petitions (urban non-pocketbook)		-0.0028
		(0.0028)
City and year fixed-effects	✓	✓
Number of Cities	308	308
Observations	1655	1655

Note: The table reports OLS results using a sample that includes provinces with low LLMB usage (Xinjiang, Qinghai, Tibet). Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.13.4 Addressing the Presence of Local Petition Platforms

Another potential concern one may have is that there are some localities that have developed their own petition platforms to receive citizen petitions and our analyses, which focus only on the centrally operated petition platform, do not capture petitions filed at those more localized sites. Although we do not see any reason as to why the patterns of participation in those more localized sites should be different from that in the LLMB (as citizens who engage in online petitions are likely to know and use both), our results

Table B9: Sub-Sample Analysis: By Region

	Subsample: Region				
	(1) East	(2) Central	(3) West	(4) Exclude Provincial Capital	(5) Exclude Autonomous City (*zhou/meng*)
Log petitions (rural pocketbook)	0.0085**	0.0065**	0.0125	0.0075***	0.0074***
	(0.0042)	(0.0025)	(0.0256)	(0.0022)	(0.0023)
Log petitions (urban pocketbook)	-0.0015	0.0034	-0.0061	0.0018	0.0017
	(0.0048)	(0.0026)	(0.0209)	(0.0024)	(0.0024)
Log petitions (rural non-pocketbook)	-0.0031	0.0011	-0.0311*	-0.0006	-0.0006
	(0.0027)	(0.0020)	(0.0155)	(0.0017)	(0.0017)
Log petitions (urban non-pocketbook)	-0.0085	-0.0014	0.0107	-0.0023	-0.0029
	(0.0053)	(0.0034)	(0.0178)	(0.0029)	(0.0029)
Year fixed-effects	✓	✓	✓	✓	✓
Province-specific trends	✓	✓	✓	✓	✓
R^2	0.07	0.04	0.73	0.05	0.04
Number of Cities	84	208	16	282	281
Observations	492	1108	55	1514	1544

Note: This table shows regression results from subsample analysis. The results suggest that the level of policy responsniveness is comparable both across geographical regions.
Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

might be affected if citizens are strategically choosing to send certain type of petitions *only* to central or local platforms. While the amount of resources and time needed for collecting systematic data on all local platforms is well beyond our current capacity, we try to address this issue by first identifying the localities that have the most active local platforms and then compare the patterns of online participation there with the rest of the country. In particular, we note that a number of local petition platforms have been awarded the China News Prize (the highest prize for journalistic/media work in China) because of their high popularity among local citizens. We compiled a full list of cities that have received such prizes as the "active local platform" group,[5] and compare LLMB participation in this group with that in other cities (i.e., no active local platform). Figure B7 displays the annual changes in average log petitions. We can see that temporal trends

[5]The list includes the following localities: (provinces) Henan, Guangdong, Liaoning, (cities) Yantai, Ningbo, Zhenjiang, and Huizhou.

of the two groups track quite closely with each other, although those with active local platforms tend to have more LLMB users during the first several years following LLMB's launch. A simple regression analysis shows that active local platform is associated with about 35% increase in LLMB petitions between 2008 and 2013. This pattern is thus consistent with our assumption that the level of LLMB usage is representative of the level of usage in other online petition platforms.

Second, we also conduct a subsample analysis to see if the patterns of responsiveness differ between localities with and without active local platforms. The results are presented in Table B9. We can see that the estimates are largely comparable between the two samples and the main estimate of interest is positive and highly significant even when we exclude those localities with more active local online platforms. Taken together, these results suggest that our findings are unlikely to be driven by the presence or absence of alternative local petition platforms.

Table B10: Sub-Sample Analysis: By Presence of Local Platforms

	Subsample: Pre-existing Local Platforms	
	(1) Yes	(2) No
Log petitions (rural pocketbook)	0.0072	0.0072***
	(0.0048)	(0.0026)
Log petitions (urban pocketbook)	0.0007	0.0029
	(0.0053)	(0.0026)
Log petitions (rural non-pocketbook)	0.0014	-0.0011
	(0.0028)	(0.0020)
Log petitions (urban non-pocketbook)	-0.0017	-0.0032
	(0.0055)	(0.0035)
Year fixed-effects	✓	✓
Province-specific trends	✓	✓
Adjusted R^2	0.09	0.04
Number of Cities	69	228
Observations	400	1224

Note: This table shows regression results from subsample analysis. The results suggest that the level of policy responsniveness is comparable between localities with and without large, active local petition platforms.
Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.14 Granger-Style Tests on Reverse Causality

We conduct several Granger-style tests to see if our results are driven by reverse causality. We run two additional regressions with the timing of the dependent variable set at t and $t-1$, respectively. If reverse causality exists, then we would expect content of GWRs to be correlated with petition intensity in current and subsequent years. However, results presented in Table B11 suggest that this is not the case: neither the current or the previous year's GWR welfare topic appears to be significantly correlated with the current year's volume of online petition. In the Online Appendix, we conduct several more extensive checks by including leads and lags of welfare topic share at both the city and the provincial level, and the main results remain largely unchanged (Table B12).

Table B11: Alternate the Timing of the Dependent Variable

	Δ Welfare Topic (10-topic) at ...			
	(1) t	(2) t	(3) $t-1$	(4) $t-1$
Log petitions	-0.0005		0.0020	
	(0.0022)		(0.0025)	
Log petitions (rural pocketbook)		-0.0036		0.0028
		(0.0023)		(0.0024)
City and year fixed-effects	✓	✓	✓	✓
Province-specific trends	✓	✓	✓	✓
R^2	0.08	0.08	0.09	0.09
Number of Cities	308	308	300	300
Observations	1607	1607	1537	1537

Note: This table shows results from altering the timing of the dependent variable. The results suggest that earlier welfare topic proportions in GWRs is not significantly associated with subsequent petition intensity.
Robust standard errors clustered at city level are reported in parentheses.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$ (two-tailed test)

In addition to the analyses presented in Table B11, we also estimate several regressions that include both the current and past welfare topic shares for provincial as well as city governments. Due to the well-known Nickell bias (Nickell, 1981), we do not include city fixed-effects in these regressions to avoid the incidental parameter problem. The results are presented in Table B12. The first two columns include lagged welfare topic share for

city GWR (t and $t-1$), and Columns 3 through 5 include the welfare topic shares in provincial GWRs at $t+1$, t, and $t-1$, respectively. Finally, the last column presents a model that includes all the preceding controls. Through all the models, we see that the coefficient estimate for type RP petitions remains quite stable. This lends us further confidence that the observed effect of online participation is unlikely to be driven solely by changes in local government policies.

Table B12: Controlling for Welfare Topic Share at the Higher Level or from Earlier Periods

	Δ Welfare Topic at $t+1$ (10-topic)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log petitions (rural pocketbook)	0.0043***	0.0053***	0.0055***	0.0076***	0.0054***	0.0039***
	(0.0013)	(0.0017)	(0.0016)	(0.0022)	(0.0016)	(0.0013)
Log petitions (urban pocketbook)	0.0015	0.0012	0.0014	0.0022	0.0014	0.0016
	(0.0015)	(0.0018)	(0.0018)	(0.0023)	(0.0018)	(0.0015)
Log petitions (rural non-pocketbook)	0.0004	-0.0004	-0.0007	-0.0005	-0.0007	0.0007
	(0.0010)	(0.0012)	(0.0012)	(0.0016)	(0.0012)	(0.0010)
Log petitions (urban non-pocketbook)	-0.0012	-0.0019	-0.0022	-0.0027	-0.0022	-0.0014
	(0.0016)	(0.0018)	(0.0018)	(0.0029)	(0.0018)	(0.0016)
Welfare topic share at t (city GWR)	-0.6814***					-0.7337***
	(0.0318)					(0.0284)
Welfare topic share at $t-1$ (city GWR)		-0.1000***				0.1243***
		(0.0290)				(0.0261)
Welfare topic share at $t+1$ (provincial GWR)			-0.0248			-0.0058
			(0.0539)			(0.0436)
Welfare topic share at t (provincial GWR)				-0.0495		-0.0656
				(0.0518)		(0.0431)
Welfare topic share at $t-1$ (provincial GWR)					0.0280	-0.0439
					(0.0470)	(0.0394)
Year fixed-effects	✓	✓	✓	✓	✓	✓
Province-specific trends	✓	✓	✓	✓	✓	✓
R^2	0.38	0.05	0.04	0.04	0.04	0.40
Number of Cities	297	295	297	297	297	295
Observations	1624	1553	1624	1624	1624	1553

Note: This table shows regression results from including a variety of controls for the welfare topic share in provincial-level GWR or that in past city-level GWRs. The results indicate that our main coefficient of interest is highly robust to the inclusion of these potential confounders, suggesting that the effect of online participation is not endogenous to governments' prior policy orientations.

Robust standard errors clustered at city level are reported in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.15 Instrumental Variables Analysis

B.15.1 Analysis Procedure

One possibility is that both policies and online participation may be driven by some unobserved time-varying events, such as a shift in societal preference for social welfare. We address this issue by looking for an instrument that is only correlated with online participation activities but uncorrelated with local demand for welfare. To do so, we exploit variations within petition topics. We construct an instrument for type RP petitions (rural, pocketbook) by calculating the average number of type UN petitions (urban, non-pocketbook) from *all other* cities in the same province. Interviews with LLMB staff suggest that the public's knowledge about the platform usually diffuses quickly within provincial boundaries, partly due to the relatively high frequency of communication and dense interpersonal networks within provinces. The average usage of the platform for non-welfare related issues should thus be a good proxy of the general popularity of the platform in a province. At the same time, the salience of rural pocketbook issues in a city should not, in theory, be directly related to the salience of urban, non-pocketbook issues in its neighbors because the two sets of issues concern different populations as well as different policy areas. While this exclusion restriction cannot be directly tested, we provide several tests in Section B.15.2 to address some potential challenges. For a city i in a province p with a total of N_p cities, the value of the instrument is calculated as:

$$\text{IV}_{ipt} \quad = \quad \log\left(\frac{\sum_{j\neq i}\text{Type } \mathbf{UN} \text{ Petitions}_{jpt}}{N_p - 1}\right).$$

Table B13 presents the IV results for GWR topic shares. Column 1 shows that the first stage correlation between the two types of petitions is indeed very strong ($F \simeq 200$), even though they cover very different issue areas. A one percent increase in type

UN petitions from neighboring cities is associated with about 0.7 percent increase in RP petitions in the city of interest. The second stage results, shown in Column 2, are consistent with the OLS results: rural, pocketbook petitions appear to have a positive and statistically significant impact on changes in welfare topic shares. The magnitude of the IV estimate is slightly smaller than the baseline, indicating that concurrent offline activities might account for some, but not all the observed effects of online petitions. In the online appendix, we further show that the IV results are robust to using instruments constructed in various different ways (Table B17 and Figure B8).

Table B13: Instrumental Variable Estimation

	Δ Welfare Topic at $t+1$ (10-topic)	
	(1) First stage	(2) Second stage
IV: average petitions in neighboring cities (urban non-pocketbook)	0.7394***	
	(0.0519)	
Log petitions (rural pocketbook)		0.0075***
		(0.0023)
City and year fixed-effects	✓	✓
First stage F	203.27	
Number of Cities	293	293
Observations	1620	1620

Note: The table reports results from instrumental variables estimation. The dependent variables are incremental increase in the share of social welfare topic in government work reports. The specification is based on Column 3 of Table 2.2.
Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

Table B14 reports the IV results on the *dibao* program, focusing on three metrics: individual coverage, family coverage, and total spending. Again, for all metrics, we see that RP petitions consistently have a positive and significant impact on the implementation of rural *dibao*: A one standard deviation increase in type RP petitions raises the program's individual coverage by about 1/7 of a standard deviation, family coverage by 1/4 of a standard deviation, and total spending by 1/5 of a standard deviation. These results are highly consistent with the OLS result reported in Section 2.5.5.

Table B14: Change in Substantive Outcomes: Rural Minimum Living Standard Guarantee Scheme (*dibao*)

	Coverage (individual)	Coverage (family)	Total spending
	(1)	(2)	(3)
	IV	IV	IV
Log petitions (rural pocketbook)	0.0552***	0.0943***	0.0682***
	(0.0193)	(0.0219)	(0.0226)
City and year fixed-effects	✓	✓	✓
First stage F	224.43	224.43	224.55
Number of Cities	300	300	300
Observations	1761	1761	1762

Note: The table reports IV regression results using several implementation statistics of the rural *dibao* program as the dependent variable. The specification is based on Table B13.
Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.15.2 Evidence on Exclusion Restriction

A key condition for our instrumental variables analysis to be valid is the satisfaction of exclusion restriction—that is, the instrument does not affect our outcome of interest in ways other than through the independent variable. In this study, it means that the average volume of urban, non-pocketbook (UN) petitions from neighboring cities does not affect the a city's welfare policy in ways other than affecting the general popularity of the LLMB in that city. There are two main threats to the exclusion restriction. First, the volume of UN petition may be correlated with certain provincial-level policies that affect all city governments' policy priorities in the same province. It is plausible, for example, that a provincial administration that cares about lower-class citizens' well-being would simultaneously encourage online petitions and ask its subordinate city governments to pay more attention to social welfare. One way to address this concern is to examine the relationship between UN petition and provincial governments' policy orientations. If this channel of influence exists, we would expect the volume of UN petitions to be first associated with how provincial governments set their own welfare policies. Table B15 reports regression results using several provincial-level welfare outcomes as the dependent variable. The first two columns use incremental changes in provincial GWR welfare

topics, and the third through fifth columns report results on fiscal spending patterns. Reassuringly, we do not find the volume of UN petitions to have significant correlation with any of the welfare outcomes. This suggests that provincial-level policy preference is unlikely to be a main confounder to our IV analysis.

Table B15: UN Petition and Provincial Welfare Policy

	Δ Provincial welfare topic at $t+1$		Provincial social expenditure at $t+1$		
	(1) 10-topic	(2) 20-topic	(3) Welfare	(4) Medical	(5) Agriculture
Log UN petitions (provincial sum)	0.0120	0.0102	-0.0203	-0.0160	-0.0382
	(0.0078)	(0.0077)	(0.0300)	(0.0298)	(0.0290)
Province and year fixed-effects	✓	✓	✓	✓	✓
R^2	0.13	0.10	0.99	0.99	0.99
Observations	144	144	144	144	144

Note: The table reports the association between the total (log) number of urban non-pocket petition in a province and provincial welfare policies. Provincial gdp and population are included as controls. Robust standard errors clustered at province level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

The second potential threat to our exclusion restriction is that changes in UN petitions in a city are directly related to welfare policy changes in that city, which would then affect policies in other cities through regional diffusion. For this mechanism to hold, we would expect to first observe a positive correlation between a city's UN petitions and its own welfare policies.[6] To test this, we regress a number of city-level welfare outcomes on a city's own level of UN petitions. The outcomes include GWR welfare topics, fiscal expenditure, and coverage of the Minimum Living Standard Guarantee Scheme (also known as *dibao*, see Section 2.5.5 for more information). As shown in Table B16, the correlation between a city's UN petitions and its own welfare outcomes are actually rather weak. This provides evidence against the diffusion hypothesis.

[6]It is worth noting that correlation does not have a transitive property (see Langford et al. (2001) for a formal proof). In other words, even the volume of UN petitions is likely to be positively correlated with pocket petitions and pocketbook petitions are positively correlated with welfare policies, this does not imply that UN petition will necessarily be positively correlated with welfare policies.

Table B16: UN Petitions and A City's Own Welfare Outcomes

	Δ Welfare Topic		Fiscal Expenditure			*Dibao* Coverage	
	(1) 10-topic	(2) 20-topic	(3) welfare	(4) medical	(5) agiculture	(6) urban	(7) rural
Log petitions (urban non-pocketbook)	0.0022	0.0024	0.0007	-0.0110	0.0114	-0.0055	-0.0144
	(0.0020)	(0.0018)	(0.0123)	(0.0112)	(0.0142)	(0.0110)	(0.0103)
Province and year fixed-effects	✓	✓	✓	✓	✓	✓	✓
R^2	0.03	0.03	0.83	0.86	0.84	0.41	0.37
Observations	1624	1624	1778	1777	1778	1762	1778

Note: The table reports the association between urban non-pocketbook petitions and a series of welfare-related outcomes at the city level. The specification is based on Column 3 of Table 2.2. Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.15.3 Robustness Checks and Placebo Tests

We use topic classifications from alternative LDA models to replicate the IV analysis. To do so, we estimate a 20-topic and a 40-topic LDA models, group the topics into the four categories (RP, UP, RN, UN) as before, and then perform the same estimation, instrumenting RP petitions with the volume of UN petitions from neighboring cities. As shown in the top and middle panels of Table B17, the main IV results remain largely unchanged. In the bottom panel of Table B17, we use the volume of UN petitions in the provincial capital as an alternative instrument, and find largely similar results.

To further address the issue that our results are driven by unobserved correlations between rural, income-related petitions and certain specific types of petitions that we use as instruments, we also run a permutation tests where we *randomly* select petition volumes for three non-rural, non-pocketbook topics to construct the instrument and re-estimate the IV regression. We repeat this exercise for 500 times and visualize the results in Figure B8. It turns out that the estimated coefficients are positive in all 500 iterations and miss the 10% threshold of significance in only 2 iterations. This permutation exercise strengthens our confidence that our IV results are not driven by the relationship between

Table B17: Alternative Instruments

	Instrumet: *Urban, Non-Pocketbook Petitions (UN) from Other Cities in the Same Province (20-Topic Model)*			
	(1) Δ Welfare topic	(2) *Dibao* Coverage (individual)	(3) *Dibao* Coverage (family)	(4) Total *dibao* spending
Log petition (rural pocketbook, 20-topic model)	0.0074***	0.0529***	0.0950***	0.0646***
	(0.0022)	(0.0187)	(0.0214)	(0.0216)
City and year fixed-effects	✓	✓	✓	✓
First stage F	231.13	245.68	245.68	245.69
Number of Cities	293	300	300	300
Observations	1620	1761	1761	1762
	Instrumet: *Urban, Non-Pocketbook Petitions (UN) from Other Cities in the Same Province (40-Topic Model)*			
	(1) Δ Welfare topic	(2) *Dibao* Coverage (individual)	(3) *Dibao* Coverage (family)	(4) Total *dibao* spending
Log petition (rural pocketbook, 40-topic model)	0.0075***	0.0525***	0.0911***	0.0642***
	(0.0022)	(0.0190)	(0.0214)	(0.0221)
City and year fixed-effects	✓	✓	✓	✓
First stage F	202.15	222.66	222.66	222.82
Number of Cities	293	300	300	300
Observations	1620	1761	1761	1762
	Instrumet: *Urban, Non-pocketbook Petitions (UN) in Provincial Capital*			
	(1) Δ Welfare topic	(2) *Dibao* Coverage (individual)	(3) *Dibao* Coverage (family)	(4) Total *dibao* spending
Log petitions (rural pocketbook)	0.0080***	0.0239	0.0443*	0.0624**
	(0.0027)	(0.0245)	(0.0248)	(0.0262)
City and year fixed-effects	✓	✓	✓	✓
First stage F	138.53	165.61	165.61	165.73
Number of Cities	293	300	300	300
Observations	1620	1761	1761	1762

Note: The table reports replication of the main topic and spending results using measures generated from a 40-topic LDA model. The specification is based on Column 3 of Table 2.2. Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

rural pocketbook petitions and any specific urban non-pocketbook topic.

Moreover, one may be worried that the diffusion of LLMB might be correlated with provincial-level shocks to welfare policies. To check whether this is the case, we conduct a placebo test using the implementation of the *dibao* program in urban areas as the alternative outcome. The results are shown in Table B18. As expected, rural pocketbook

petitions have little effect on the distribution of urban *dibao*. This gives us confidence that our finding is not driven by unobserved correlations between platform diffusion and governments' preferences for welfare policies.

Table B18: Placebo using Urban Dibao as Outcome

	No. individuals covered (log)	Total spending (log)	Per capita spending (log)
	(1) IV	(2) IV	(3) IV
Log petitions (rural pocketbook)	0.0012	-0.0308	-0.0163
	(0.0182)	(0.0197)	(0.0110)
City and year fixed-effects	✓	✓	✓
First stage F	229.54	229.54	229.54
Number of Cities	302	302	302
Observations	1777	1777	1777

Note: The table reports a placebo test using urban *dibao* spending as the outcome. The result suggests that rural pocketbook petitions have no effect on the implementation of urban *dibao*. The specification is based on Column 3 of Table 2.2. Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.16 Extensions

B.16.1 Effects of Online Petitions on Spending Patterns

Table B19: Effects of Online Petitions on Fiscal Expenditures

	Growth in Expenditure			
	(1) social	(2) medical	(3) education	(4) science
Log petitions	0.0570	-0.2215	0.0326	-0.0012
	(0.0452)	(0.2124)	(0.0227)	(0.0319)
Year and city fixed-effects	✓	✓	✓	✓
Economic controls	✓	✓	✓	✓
Leadership controls	✓	✓	✓	✓
R^2	0.16	0.02	0.13	0.04
Number of Cities	303	303	303	280
Observations	1773	1777	1762	1637

Note: The table reports the estimated effects of online petition on government expenditure in several areas. The specification is based on Column 3 of Table 2.2. Robust standard errors clustered at city level are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)

B.16.2 Effects of Participation over Time

We also investigate the possibility that our results were driven by rare events that happened in a few special years. To do so, we estimate a model where the participation variable is interacted with a set of year dummies. It is clear that the effect is in fact quite stable over time. All estimates are statistically significant at 10% level or higher with the exception of 2011. This pattern lends us further confidence that the result is not driven by the choice of sample period.

Figure B1: A Snapshot of the LLMB

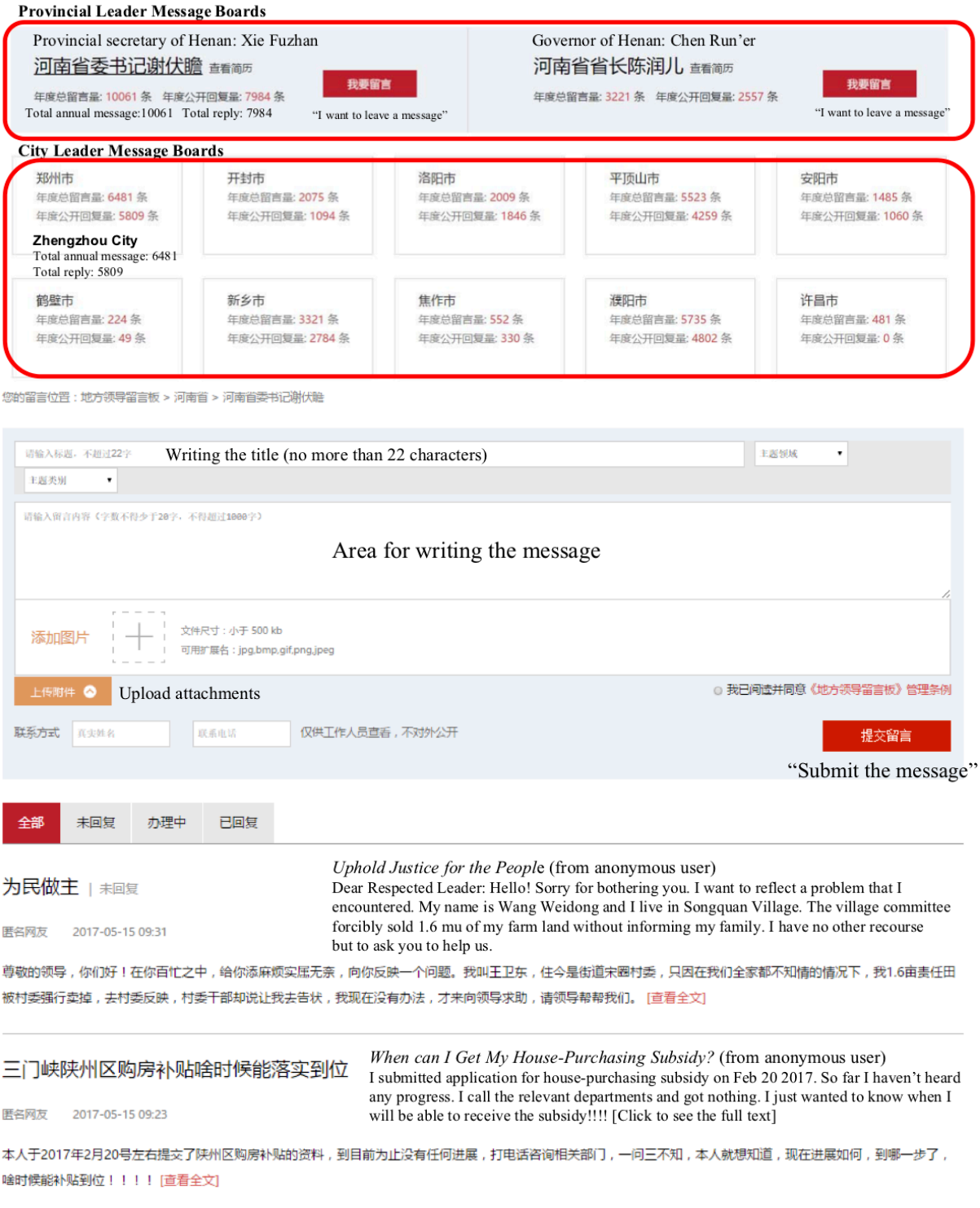


Figure B2: Posterior Distribution of Topic Proportions in the Whole Corpus

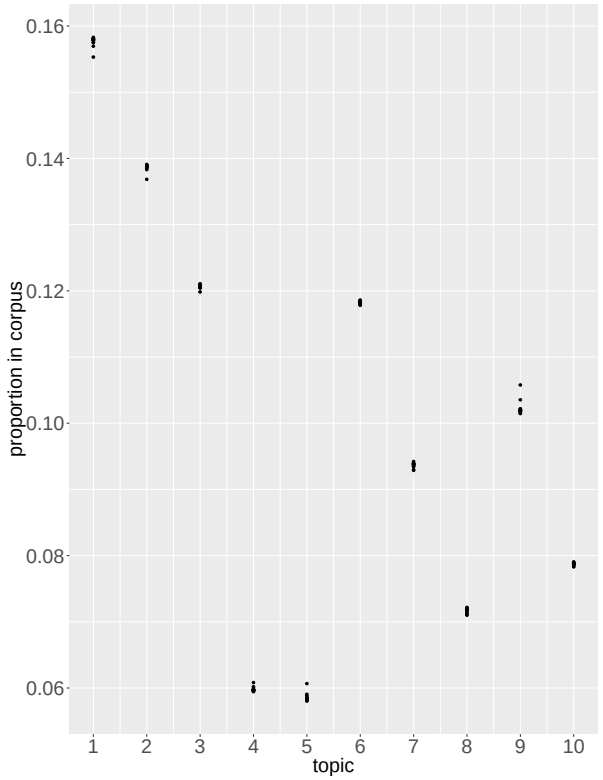
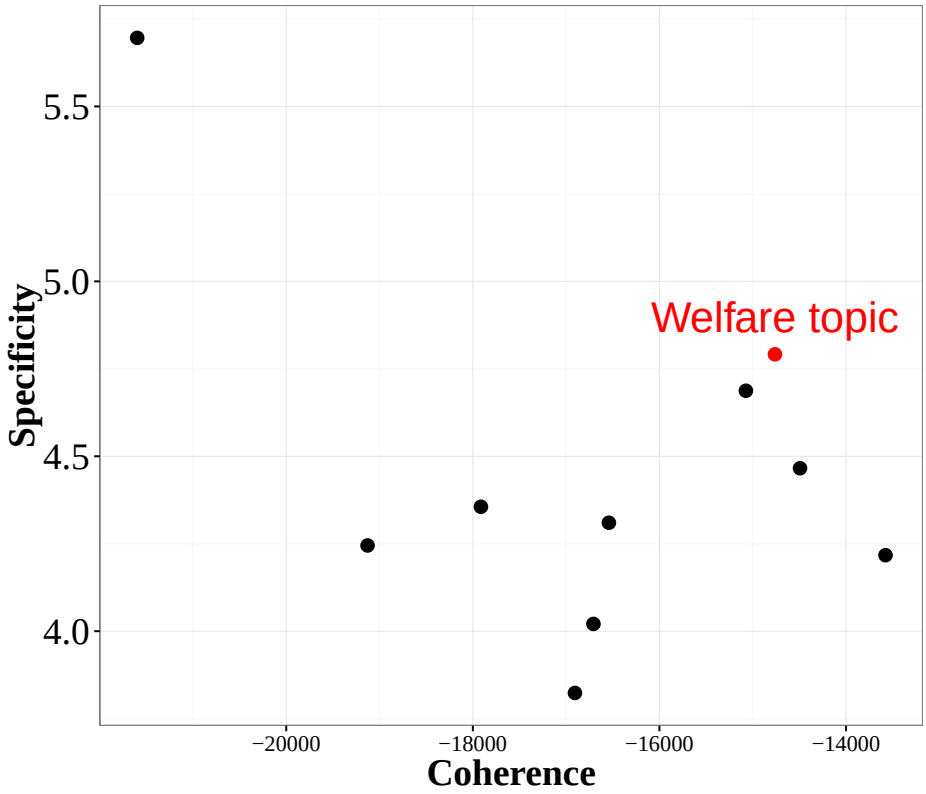


Figure B3: Topic Quality: Coherence vs. Specificity



Note: This figure illustrates the quality of the welfare topic in terms of both specificity and coherence, in comparison with other nine topics.

Figure B4: Average Number of Petitions per Capita, 2008-2013

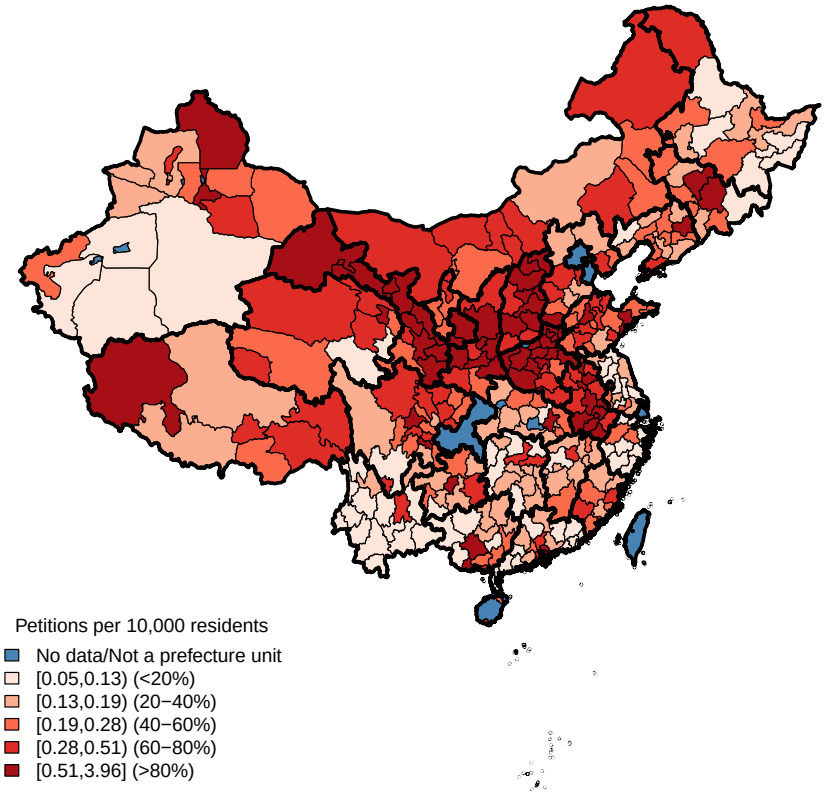
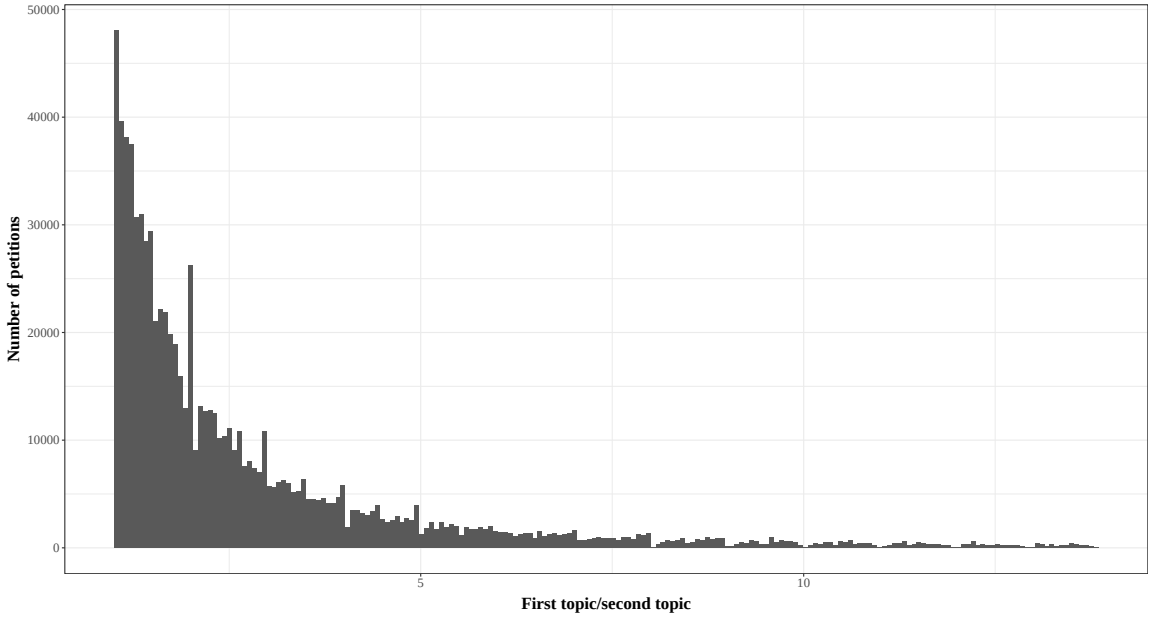


Figure B5: Clear Dominant Topic in Petitions



Note: This figure plots the distribution of largest topic proportion over second-largest topic proportion in petitions. The 5% right tail is truncated.

Figure B7: Level of LLMB Participation by Presence of Active Local Petition Platform

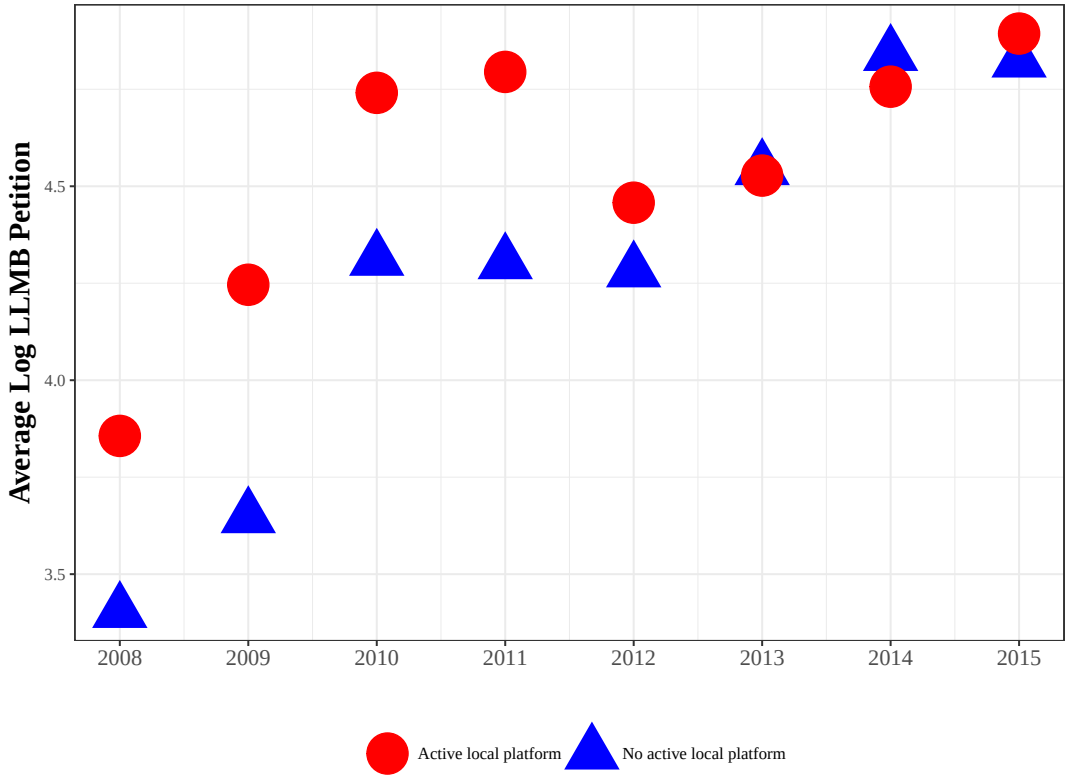
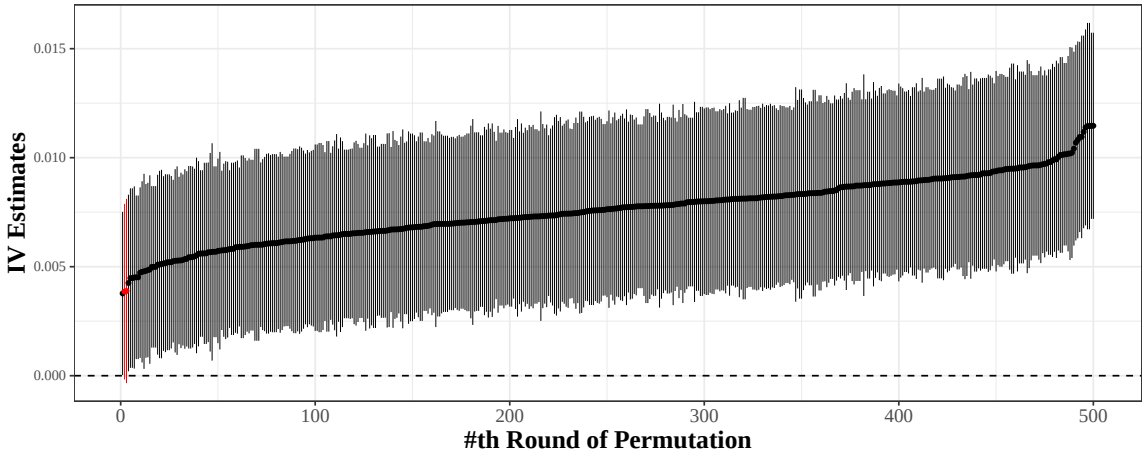
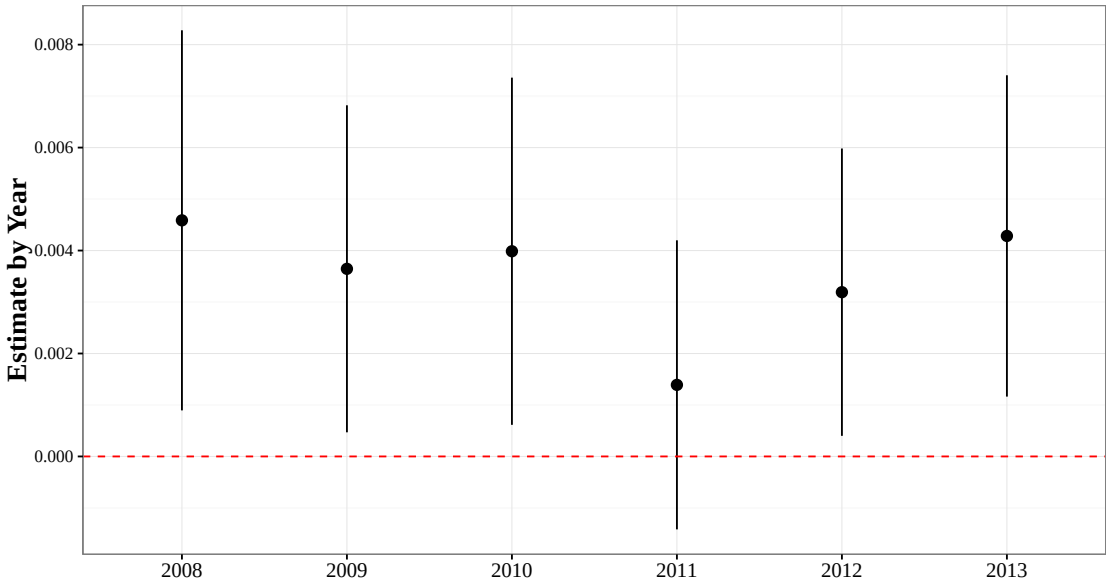


Figure B8: Permutation of Topics Used for Instrument



Note: This figure visually presents the results from 500 permutations of the instrument. Only 2 out of 500 permutations turn out to be statistically insignificant at 10% level (highlighted in red)

Figure B9: Effect of Online Participation by Year



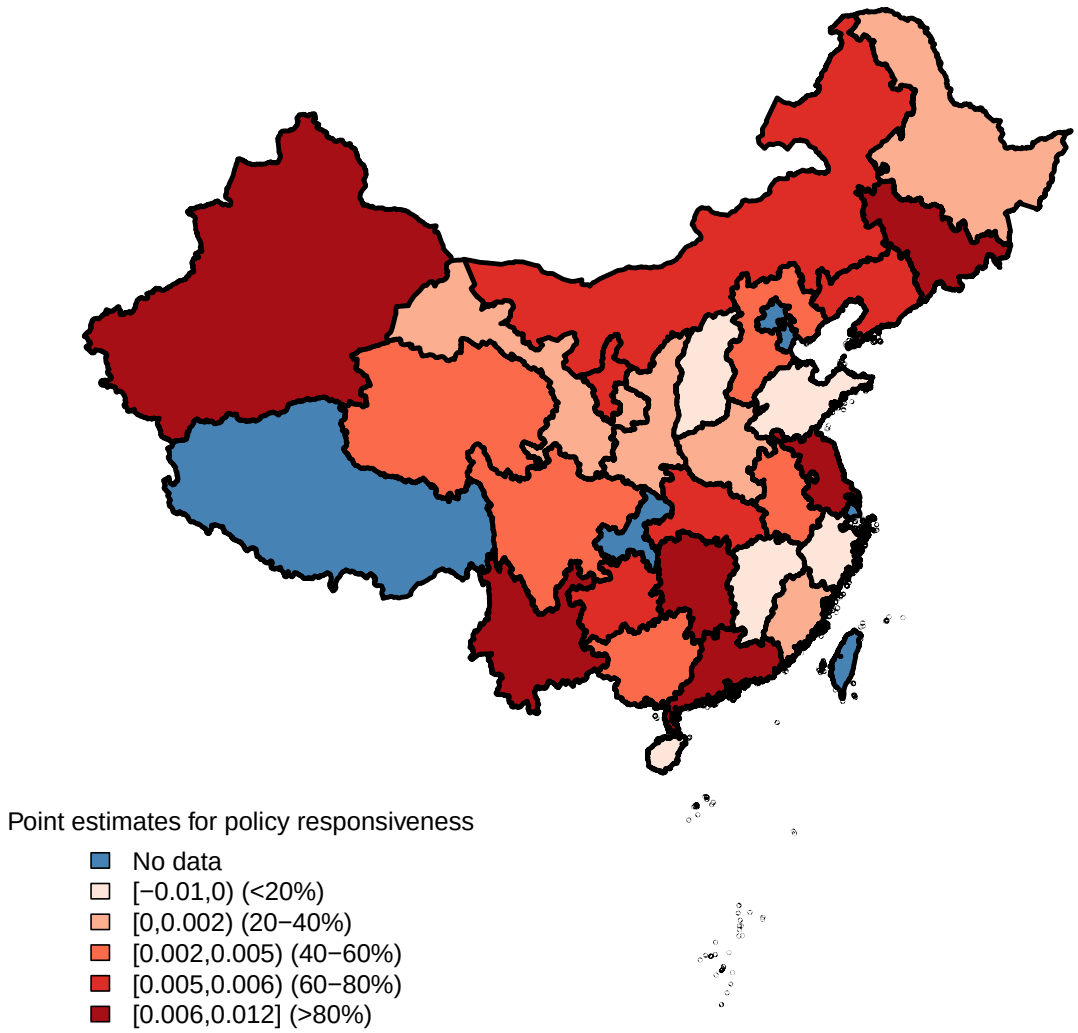
B.16.3 Geographic Variation in Responsiveness

We also estimate province-specific policy responsiveness using the following equation:

$$\Delta\text{Welfare Topic Share}_{it+1} = \alpha\text{Log Petitions}_{it} + \delta^{p}\text{Log Petitions}_{it} \times \text{Province}_{p} + \mathbf{X}_{it}\boldsymbol{\beta} + \phi_i + \tau_t + \epsilon_{it}.$$

The level of policy responsiveness for province p is thus $\alpha + \delta^{p}$. We plot this quantity in the following map.

Figure B10: The Geography of Policy Responsiveness



B.17 Fieldwork Evidence on the Petition-Policy Link

In this section, we draw on materials from our fieldwork to flesh out the process by which online petitions affect government policy making. Between 2013 and 2017, we conducted several dozens of interviews with central and local government officials and staff at the People Online (the company that manages LLMB) to learn about how Chinese government handles online petitions. According to our interviews, most local governments today have set up specialized agencies dedicated to monitoring social media sites and online petition platforms. These agencies, usually under names such as Public Opinion Office and Social Condition and Mass Opinion Office, are responsible for collecting information about online public opinion from popular websites and reporting it to key local decision makers (e.g., party secretary, mayor, or members of the local party standing committee). The reporting takes several forms: The first are daily, weekly, or monthly briefings that summarize trending issues in online discussion. These reports usually cover a wide range of policy or non-policy topics and are intended to keep local leaders updated about changes in the focus of public attention. A second type of reports are in-depth investigation on public reactions to specific accidents or events, such as scandals, collective protests, or safety accidents. This type of reports usually combines online public opinion information with on-ground research to produce actionable plans for local authority to manage and direct public sentiment. Finally, a third type of reports are policy recommendations based on public opinion in a specific policy domain (e.g., education, health care, or social welfare). When drafting guiding policy documents, such as the Government Work Reports, local governments will consult a wide range of agencies and actors, and input from online public opinion has become increasingly important in recent years as the influence of online platforms grow.[7]

[7]The central government, for example, initiated a campaign entitled "I spoke to the Premier" to encourage netizen input to the drafting of the Central Government Work Report in 2014. Many local

Our fieldwork also suggests that the specific ways by which local governments handle online petitions vary greatly across localities. Usually the handling of online petitions is led by one the following agencies: the general office of party committee, the government general office, the cyber administration office, the publicity bureau, the supervision department, or the letters and visits bureau. The most popular choice is to manage online petitions through the general office of the local party committee; this practice are adopted by eight provinces, including Zhejiang, Anhui, and Gansu. Another popular model uses the government general office as the leading management agency; Tianjin, Shanghai, and Liaoning utilize this mode. In some other provinces, such as Beijing and Shandong, the cyber administration office is in charge of monitoring petitions and issuing replies. In Hunan, online petitions are managed by the supervision department. In Fujian, the letters and visits bureau is authorized to manage the online petitions. Shanxi, moreover, created a new omnibus agency called public opinion office to carry out all tasks related to online public opinion management.

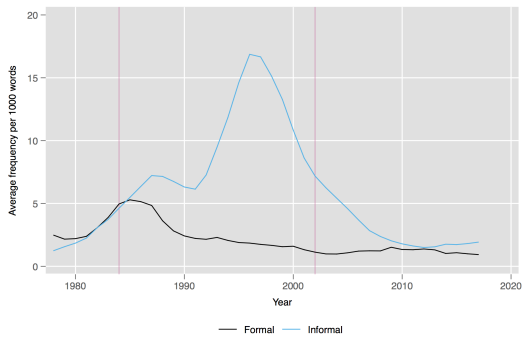
governments have also made their own initiatives. See https://goo.gl/d1oDQP; https://goo.gl/qwS8SG; https://goo.gl/nyFHAC.

Appendix C

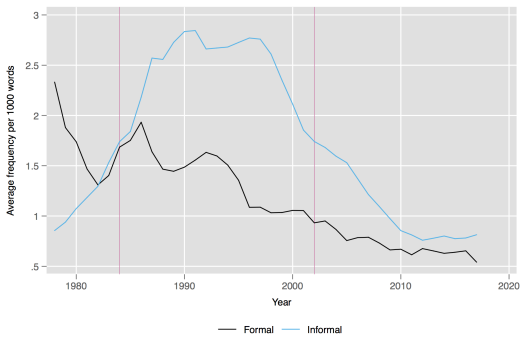
Appendix to Chapter 3

C.1 Aggregate Trends for Washington Consensus Components

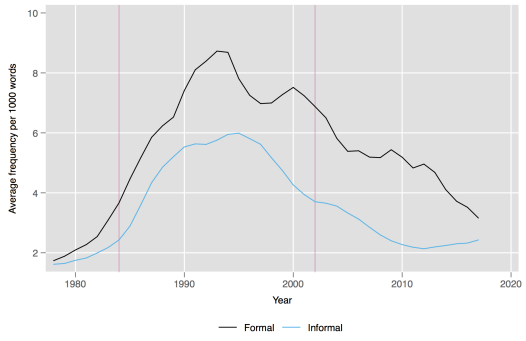
Figure C1 reports the share of words similar to each of the ten Washington Consensus components in the corpus, where similarity is determined by word embeddings.



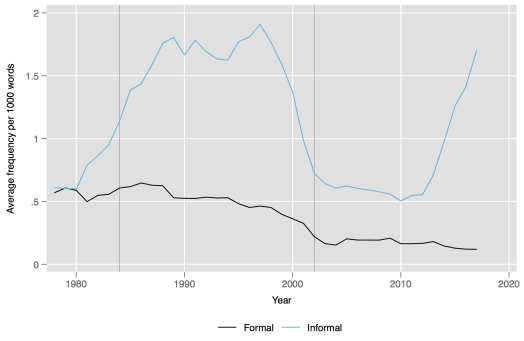
(a) Taxation



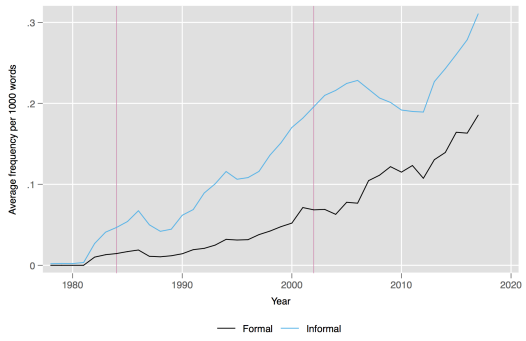
(b) Trade



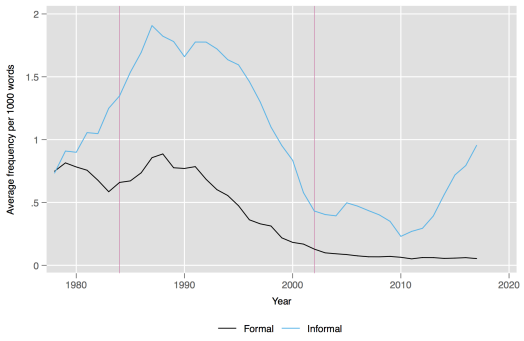
(c) Property rights



(d) Interest rate

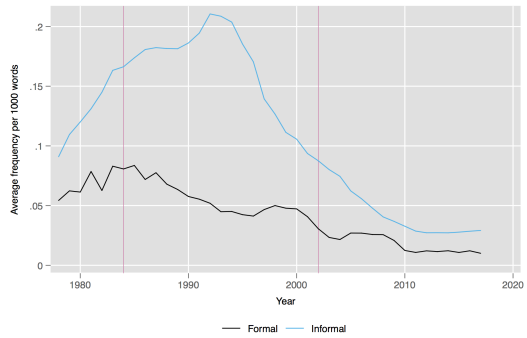


(e) Deregulation

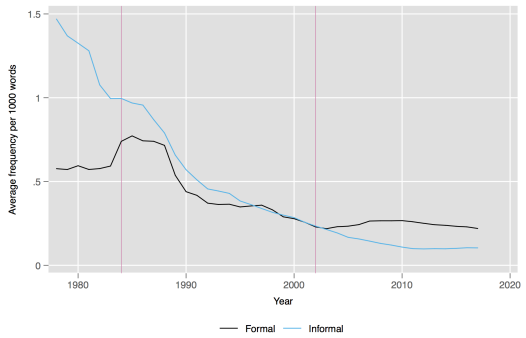


(f) Exchange rate

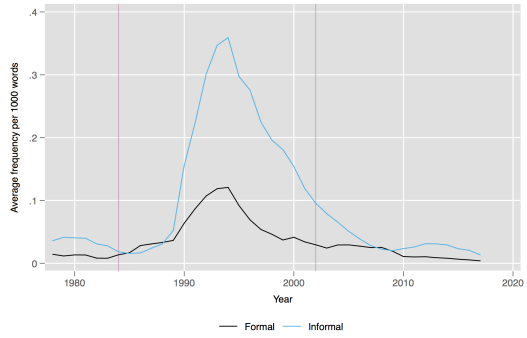
Figure C1: Washington Consensus Components in Formal vs. Informal Laws



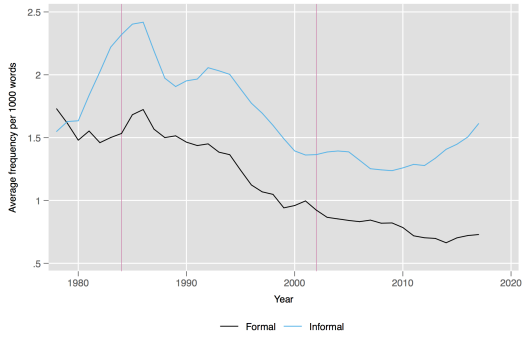
(g) Fiscal deficit



(h) Fiscal expenditure



(i) Privatization



(j) FDI

Figure C1: Washington Consensus Components in Formal vs. Informal Laws

Notes: These figures report the share of words similar to each component of Washington Consensus in legal documents enacted each year, broken down into formal and informal laws. Words with cosine similarity (to anchor words) higher than 0.4 are counted. Five-year moving averages are plotted.